Horia D. PITARIU Monica ALBU

PSIHOLOGIA PERSONALULUI:

I. MĂSURAREA ȘI INTERPRETAREA DIFERENȚELOR INDIVIDUALE

Editura Presa Universitară Clujeană
Cluj-Napoca
1996

COLECȚIA UNIVERSITARIA 4
Cursuri universitare pentru studenți

PSIHOLOGIA PERSONALULUI: I. MĂSURAREA ȘI INTERPRETAREA DIFERENȚELOR INDIVIDUALE

Horia D. PITARIU, Monica ALBU

Tehnoredactare computerizată: A. Zaharie

Editura Presa Universitară Clujeană
Director: Andrian Marcus
M. Kogălniceanu nr. 1
3400 CLUJ-NAPOCA
ROMÂNIA
Tel.: 40-064-194315
Fax: 40-064-191906

I.S.B.N. 973-9261-19-1

CUPRINS

PREFAȚĂ

Domeniul managementului resurselor umane a primit în zilele noastre o importanță covârșitoare, fiind considerat ca una din problemele cheie ale funcționării inteligente a oricărei organizații, indiferent că este vorba de organizații orientate pe business sau industriale, educaționale, servicii, militare, cu caracter social etc. Directorii de resurse umane în calitatea lor de manageri au responsabilități care urmează îndeaproape pe cele ale președintelui unei companii sau director general. Succesul unei organizații depinde în foarte mare măsură de calitatea factorului uman sau de gestionarea științifică a personalului. În cadrul Departamentului de Resurse Umane, găsim Compartimentul de Psihologia Personalului ale cărui atribuții includ problemele de selecție a personalului, formarea și perfecționarea forței de muncă, promovarea și evaluarea performanțelor profesionale etc.

Lucrarea prezentă acoperă un spectru restrâns, acela al utilizării testelor psihologice sau partea de psihodiagnoză industrială. Ea este destinată psihologilor industriali și celor care se pregătesc pentru aplicarea psihologiei la nivel organizațional. Am avut în vedere în primul rând calitatea științifică a cunoștințelor incluse, faptul că utilizarea instrumentarului psihodiagnostic solicită o pregătire de excepție dacă se dorește într-adevăr să fie eficientă, competitivă. Decizia luată pe bază de teste psihologice poate marca întreaga viață a unui individ, poate lăsa urme de lungă durată, așa că orice eșec la acest nivel poate avea consecințe serioase. Apoi, intenția noastră a fost și aceea de a dezvălui complexitatea unei problematici în care diletantismul este în floare, unde încă se consideră că cine are un test poate să-l și folosească. Firmele de consultanță, din nefericire și unele instituții de prestigiu, pretind că fac selecția managerilor sau evaluarea resurselor umane, dar metodologia lor este una desuetă, adesea bazată pe un instrumentar preluat necritic și nesupus procedurilor de adaptare la obiectivele pretinse sau la mediul nostru cultural. Firește credibilitatea deciziilor este pe măsură.

Managementul resurselor umane presupune o activitate vastă și complexă la care își aduc aportul specialiști din diferite domenii ale științei. Intenția noastră este să prezentăm sub forma unui ciclu problemele psihologice din această sferă. În ceea ce privește lucrarea de față, prima din ciclul amintit, ea conține cunoștințele teoretice fundamentale pe care orice psiholog este dator să le cunoască și să le aplice în practica sa. În capitolul 1 este discutat raportul dintre psihologia aplicată și managementul resurselor umane, perspectivele acestor preocupări în contextul nivelului actual de dezvoltare al științei și tehnicii. Capitolul 2 ne prezintă unele noțiuni fundamentale despre măsurarea specifică științelor sociale, accentul căzând pe esența și specificitatea măsurării psihologice. Totodată este detaliată noțiunea de test psihologic. Calităților testelor psihologice, fidelitatea și validitatea, le este dedicat capitolul al 3-lea. Considerăm acest capitol de o importanță majoră deoarece el dă consistență oricăror intervenții psihologice în care sunt utilizate teste psihologice. Mai mult, au fost introduse, ca de altfel pe tot parcursul lucrării, problemele noi ale psihodiagnozei, concepțiile prezente care adesea au schimbat radical atât conținutul cât și contextul contemporan al psihodiagnozei. Abordarea caracteristică școlii cognitiviste se resimte cu mai multă pregnanță în capitolul al 4-lea care discută implicațiile analizei de itemi asupra creșterii, în general, a puterii predictive a testelor psihologice.

Documentarea necesară realizării acestei lucrări a fost posibilă datorită sponsorizării autorilor ei de către International Research Excenge Office (IREX) (1992), Deutsche Forschungs Gesellschaft (DFG) (1995) și Comunitatea Europeană (Grant CIPA3510ct924603). Grantul acordat de Academia Română (1995) a permis obținerea unor cărți utile.

Autorii

CAPITOLUL 1

PSIHOLOGIA APLICATĂ ȘI MANAGEMENTUL RESURSELOR UMANE

Organizațiile, munca și psihologia aplicată - precizări și delimitări

Psihologia aplicată în diferite domenii ale vieții sociale, în școală, industrie, clinică etc., are ca obiectiv fundamental studierea omului, preocuparea pentru calitatea vieții, reducerea stresului cotidian - indiferent că este vorba de stresul școlar, organizațional, etc.

Psihologia aplicată la nivel organizațional este preocupată în mod nemijlocit de a face organizațiile mai eficiente și mai atrăgătoare ca locuri de muncă. Ea are în vedere omul în contextul impactului cu mediul muncii.

Managementul resurselor umane sau managementul modern al personalului, este un domeniu care se referă la filosofia, politica, procedurile și practicile legate de conducerea oamenilor din organizații (French, 1990). În mod obișnuit sunt incluse aici mai multe tipuri de activități ca: planificarea resurselor umane, proiectarea și reproiectarea muncii, sistemul de salarizare și recompensă, modernizarea organizațională, psihologia inginerească, psihologia personalului, medicina muncii, relațiile umane etc. Psihologia personalului este deci un subdomeniu al managementului resurselor umane, în ea fiind incluse activități ca recrutarea, selecția și repartizarea personalului, reorientarea profesională, evaluarea performanțelor profesionale și formarea profesională. Desigur, psihologia personalului este o parte componentă a psihologiei muncii și organizaționale (psihologia M/O), al cărei obiect general îl formează studiul comportamentului uman în contextul muncii. Un

lucru este însă clar, ***psihologia personalului este o ramură a psihologiei M/O care intră în atribuțiile psihologilor industriali.***

În zilele noastre, problemele de personal au devenit mult mai complexe și mai diversificate datorită unui complex de factori de natură politică, socială, economică și tehnică. Într-o companie, de orice natură ar fi ea, *Directorul de Resurse Umane*. este unul din factorii de decizie importanți. Responsabilitățile sale sunt mari, printre altele el fiind răspunzător de recrutarea, selecția și instruirea personalului, evaluarea competenței profesionale a salariaților, promovarea personalului, aplicarea corectă a legislației muncii la problemele de personal etc. Acesta este motivul pentru care un manager care răspunde de sectorul resurselor umane trebuie să fie un bun cunoscător al teoriilor organizaționale, să fie conștient de implicațiile pe care le au deciziile sale, erorile din această sferă având repercusiuni dramatice. Costurile, eficiența personalului, frustrările, nemulțumirile din rândul salariaților sunt greu de estimat, consecințele fiind extrem de scumpe. Se apreciază că activitatea de recrutare, selecție și instruire poate costa o companie, în general, în jur de 30.000 dolari. Iată de ce problemelor de personal li se acordă în zilele noastre o așa de mare importanță.

În Figura 1-1 prezentăm o schemă organizatorică a compartimentului de resurse umane pentru o întreprindere mică și mijlocie. Figura 1-2 ilustrează structura Direcției de Resurse Umane a unei întreprinderi mari.

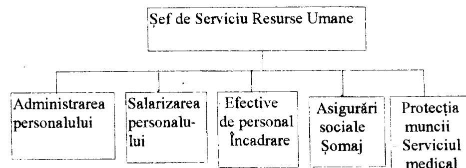


Figura 1-1. Structura Compartimentului de Resurse Umane al unei companii de dimensiuni mici și mijlocii (Lefter & Manolescu, 1995)

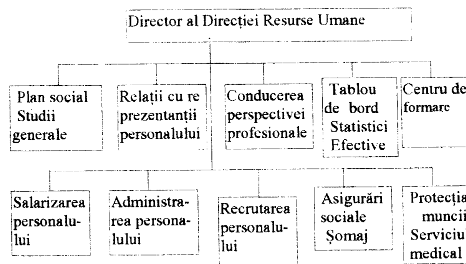


Figura 1-2. Structura Direcției de Resurse Umane a unei companii mari (Lefter & Manolescu, 1995)

Lefter și Manolescu (1995) ilustrează un mod de cuantificare a activităților de selecție, încadrare și integrare a personalului.

Aceste cheltuieli includ:

(I) Costurile de intrare în întreprindere a unei persoane:

- salariile și celelalte cheltuieli legate de personalul implicat în activitatea de angajare (personalul operațional, operatorii de interviu, psihologi, medici,.
- cheltuieli de publicitate și alte tipuri de cheltuieli: anunțuri, onorariile cabinetelor specializate în recrutarea forței de muncă, materiale de birou, consumabile etc.

(II) Costurile de integrare a salariatului. Acestea se distribuie de-a lungul a trei faze:

(1) Faza de informare (A): este faza de familiarizare cu postul de muncă - studiul documentelor postului, cunoașterea colegilor, a colaboratorilor, a șefilor și a subordonaților. Pentru această fază sunt necesare câteva zile, timp în care eficiența angajatului este foarte scăzută.

(2) Faza de însușire a meseriei (B): salariatul începe să lucreze, învață să manevreze utilajul la care muncește, adică debutează în activitatea pentru care a fost angajat. Uneori, în această fază, noul angajat este pus în situația să urmeze un curs de formare profesională sau de reîmprospătare a cunoștințelor. Lucrând, angajatul devine din ce în ce mai eficient, numărul erorilor scăzând treptat.

(3) Faza de contribuție personală în întreprindere (C): această fază se caracterizează prin faptul că salariatul este stăpân pe "meseria" sa și contribuie din plin la realizarea obiectivului de eficiență a postului său.

Considerând 100% nivelul normal de eficiență a salariatului (standardul impus de întreprindere), eficiența sa în cele trei faze se poate reprezenta grafic ca în Figura 1-3. Menționăm că noneficiența trebuie recuperată de la angajat din salariul său lunar, proporțional cu evoluția sa de-a lungul celor trei faze.

Pentru a descrie raportul dintre costurile de recrutare/selecție și încadrare și cele de integrare, se calculează un coeficient al costurilor care revin noului salariat (Cns)(*)

Costurile integrării au în vedere costurile noneficienței din perioada de integrare, corespunzător celor trei faze amintite. Acest coeficient (Cns) ia valori diferite: pentru muncitori el este cuprins între 2.5 - 8; pentru funcționari el este cuprins între 2 - 4.

Din punctul de vedere al managementului resurselor umane trebuie să se urmărească adoptarea unei politici de personal care să ducă la scăderea mărimii coeficientului respectiv.

(*) Cns = costurile integrării / (costurile recrutării/selecției + costurile de încadrare)

~~Pentru a descrie raportul dintre costurile de recrutare/selecție și încadrare și cele de integrare, se calculează un coeficient al costurilor care revin noului salariat (Cns)~~:

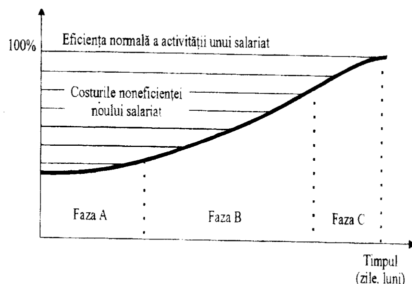


Figura 1-3. Evoluţia eficienţei activităţii unui nou angajat de-a lungul celor trei faze de integrare profesională (Lefter & Manolescu, 1995).

Organizaţia

Am utilizat adesea termenul de "organizaţie" şi mai puţin pe cel de "companie" sau "întreprindere". În primul rând, organizaţia acoperă o sferă mai largă care include şi companiile sau întreprinderile, indiferent de natura lor. De-a lungul vieţii fiecărui individ, acesta este inclus sau este în contact cu diferite organizări sociale sau economice ori, simplu, cu organizaţii umane: şcoală, organizaţii militare, cluburi, întreprinderi, spitale etc. În cadrul acestora fiecare dintre noi interacţionăm cu alţi indivizi, cu numeroase obiective etc. Distingem însă câteva puncte comune care se potrivesc la toate organizaţiile.

O organizaţie este o unitate socială proiectată deliberat să îndeplinească un obiectiv sau câteva obiective specifice (de exemplu, şcoala, o companie de proiectat software, o organizaţie militară, un spital etc.). Orice organizaţie reprezintă un sistem deschis compus

din trei elemente fundamentale: intrări, comunicaţii/procesare şi ieşiri. Intrările (materia primă) reprezintă importurile din mediul extern. În organizaţie produsele intrate sunt prelucrate, transformate, modificate (de exemplu, într-un şantier naval, în fiecare zi sunt prelucrate tone de oţel). Ieşirea constă în exportarea în mediul extern a produsului finit. În tot acest sistem, la toate nivelurile ne confruntăm cu indivizi izolaţi sau colectivităţi umane care sunt ingredientele fundamentale ale tuturor organizaţiilor.

Această lucrare tratează un grupaj de probleme aferente oamenilor ca membri şi resurse ale organizaţiilor, arată modul în care psihologia personalului poate să-şi aducă contribuţia la sprijinirea organizaţiilor ca să devină mai înţelepte, să acţioneze mai inteligent şi mai uman în utilizarea sau gestionarea resurselor umane.

Psihologia personalului se ocupă de diferenţele individuale legate de comportamentul şi performanţa în muncă şi cu măsurarea şi predicţia acestor diferenţe.

Diferenţe specifice lumii profesiunilor

Lumea profesiunilor se caracterizează prin varietate şi dinamică. Cunoscutul *Dictionary of Occupational Titles* (DOT) publicat de US Department of Labor în anul 1977, descrie 12860 profesii existente în SUA. Ministerul Muncii din România va realiza *Clasificarea Ocupaţiilor din România* (COR) care totalizează aproximativ 3000 de profesii existente în ţara noastră. Profesiile nu sunt entităţi stabile, ci sunt în continuă modificare. Oţelarul topitor de azi diferă foarte mult de cel de acum 30 de ani, inginerul chimist de azi şi-a schimbat mult profilul profesional faţă de cel de acum 50 de ani. Automatizarea şi computerizarea au modificat mult conţinutul multor profesii. Dinamica profesiunilor este o realitate, unele profesii dispar, altele apar. Chiar cererea de personal variază şi ea în timp. În Figura 1-4 prezentăm o predicţie făcută de *The US Department of Labor*

(1988) privitor la tendința de creștere a numărului locurilor de muncă în diferite grupuri profesionale.

Confruntarea cu această gamă largă de profesii aflate într-o continuă mișcare și cu evantaiul larg de patterne comportamentale și de valori, cu aspirațiile, interesele și abilitățile diferite ale oamenilor, subliniază încă o dată importanța utilizării eficiente a programelor de resurse umane.

Diferențe în performanță. Am arătat că oamenii constituie una din sursele importante de investiții din partea companiilor; costurile de recrutare, selecție și formare sunt foarte mari în contextul industriei automatizate. Sub aspect psihologic știm însă că oamenii sunt diferiți. Ei se deosebesc unii de alții prin înălțime, greutate și alte dimensiuni fizice, la fel ca și prin aptitudini, deprinderi, interese etc. De pildă, să observăm cum lucrează un grup de operatoare care introduc date în calculator.

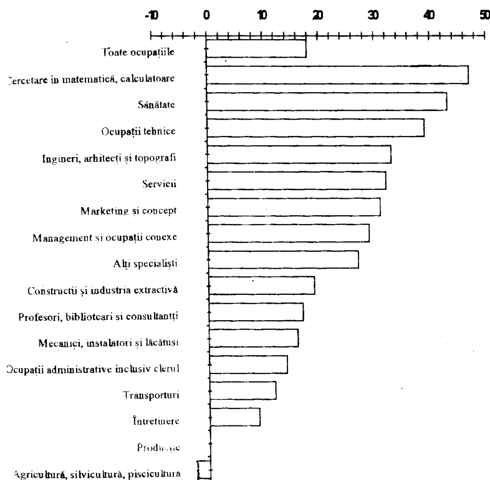


Figura 1-4. Estimarea creșterii numărului de angajați în diferite profesii până în anul 2000 (*The US Department of Labor*, 1988).

Vom vedea că există stiluri diferite de lucru: una urmărește tastatura calculatorului ca să nu facă erori, alta tastează fără să mai necesite un continuu autocontrol, una are un ritm de lucru alert, în timp ce alta

este mai lentă etc. Aceste comportamente duc la variații în performanță. Aici intră în joc numeroși factori cum sunt aptitudinile, nivelul de dezvoltare al deprinderilor de lucru, factorii temperamentali, motivaționali etc. S-a constatat că motivația joacă un rol major în performanța profesională. De exemplu, două grupe egale ca inteligență, dar diferite ca motivație, vor fi diferite și ca performanță. Grupul cu motivație mai mare va avea un succes mai mare într-o sarcină de rezolvarea de probleme. Rezultate similare s-au găsit când indivizii efectuează sarcini simple perceptual-motorii (Eysenck, 1963). Pentru activități de muncă simple, motivația poate fi cea mai importantă variabilă determinantă a performanței angajaților, mai ales în unele situații în care aptitudinile și pregătirea cerute sunt minime.

Munca și societatea

În societatea modernă, munca a devenit activitatea centrală a vieții indivizilor. Aproape o treime din viața sa, omul și-o petrece muncind. Dar, din anii 70 și apoi 80, o serie de cercetări au arătat că societatea contemporană se confruntă cu o nouă psihologie a muncii. Aceasta are trei componente fundamentale:

(1) Creșterea importanței timpului liber, ca o consecință a trebuințelor de autorealizare.

(2) Accentuarea semnificației simbolice a muncii plătite, în țările dezvoltate (mai ales pentru femei). Munca în sine este privită mai mult ca un simbol al oportunității de a aparține ca membru al unei societăți, un simbol al autoaprecierii.

(3) Tendința ca munca să devină mai puțin depersonalizată. "Recunoașterea ca individ" este considerată ca cel mai important aspect al muncii. La fel, oportunitatea "să fii în prezența unor oameni plăcuți, cu care să-ți placă să lucrezi", este un deziderat al muncitorului de azi.

Pentru majoritatea oamenilor, aceste cerințe vin în preîntâmpinarea dorinței ca munca însăși să fie interesantă și nonrutinieră.

Lecția fundamentală învățată de către societatea modernă din experiența marilor companii poate fi rezumată concis:

> *Tratează oamenii ca adulți. Tratează-i ca parteneri; tratează-i cu demnitate; tratează-i cu respect. Tratează-i - nu ca și consumatori de capital și automatizare - ci ca sursă primară de câștiguri productive ... Dacă dorești productivitate și recompensă financiară, trebuie să-ți tratezi angajații ca cea mai importantă investiție.*

Tema enunțată revine frecvent ca un laitmotiv în studiile despre satisfacția profesională. O componentă importantă a motivației în muncă este "autonomia muncii", măsura în care angajații simt că ei pot lua propriile lor decizii și pot să influențeze într-un fel sau altul ce se întâmplă la nivelul postului lor de muncă. Această precizare poate explica de ce, în societățile avansate, salariul nu reprezintă întotdeauna sursa principală de satisfacție. Cu alte cuvinte, cu cât autonomia muncii este mai mare, cu atât mai mulțumiți sunt salariații de ceea ce fac.

Dacă ne punem întrebarea de ce muncesc oamenii, răspunsul nu este simplu, el comportă o abordare complexă cu multe nuanțe. Succesul în viață, sensul progresului, al dezvoltării personale, realizarea profesională, siguranța și prestigiul, toate aceste realități se leagă de munca pe care un individ o prestează. Standardul de viață al unui individ este determinat într-o măsură semnificativă și de munca sa. Înseamnă că efectele muncii nu pot fi delimitate de cele opt ore pe care un individ le petrece muncind într-un anumit loc de muncă, ele se extind asupra întregii vieți a acestuia.

Perspectivele psihologiei personalului

Oamenii au constituit subiectul de studiu predilect al psihologilor, iar comportamentul oamenilor în contextul muncii a făcut obiectul de studiu al psihologilor muncii sau organizaționali. Psihologia personalului ca subdomeniu al psihologiei M/O este o disciplină aplicată la studiul diferențelor individuale în procesul complex al muncii. Printre domeniile sau ramurile de interes ale psihologiei personalului menționăm: analiza și evaluarea muncii, recrutarea, preselecția, selecția și repartiția personalului, instruirea și perfecționarea profesională a personalului, aprecierea performanțelor profesionale etc.

Psihologia personalului este un domeniu care în prezent cunoaște o revigorare, o dată cu sporirea ponderii automatizării proceselor industriale. Problemele majore cu care se confruntă psihologia personalului în prezent privesc: legalitatea deciziilor de personal dintr-o companie; aspectele sociale ale muncii; problematica economică; politica organizațională etc. Analizând problemele psihologiei muncii și organizaționale din Europa de Est, Roe (1995) subliniază câteva aspecte specifice acestei discipline în perioada de tranziție socială pe care o traversează țările respective. Astfel, o caracteristică importantă a economiei de tranziție este marea privatizare, spargerea marilor companii în companii mici. Sub aspect individual sau al grupului, prin abandonarea structurilor organizaționale stabile și implicit al patternelor comportamentale relativ fixe, se vor crea premisele unor situații incerte cu tot cortegiul nefast al acestora asupra oamenilor. Penetrarea treptată a economiei de piață în țările amintite va pretinde o concentrare a activității psihologice pe formarea de noi patterne de management și conducere, obiectivele manageriale se vor schimba, la fel stilul de conducere, care va adopta valențe noi, democratice, cu eficiență sporită. Asistența psihologică a șomerilor constituie o preocupare fundamentală a psihologilor, cu un spectru de activități extrem de larg, de la orientare și reorientare profesională, la selecție și organizarea de sisteme de formare profesională. Sfera de acțiune a psihologilor se va lărgi considerabil prin extinderea domeniilor de acțiune la aplicații până nu de mult neglijate, ca: motivația pentru muncă, stresul ocupațional, calitatea vieții profesionale/ocupaționale, dezvoltarea organizațională, evaluarea performanțelor etc.

Referitor la problemele juridice care pot să apară legat de intervențiile psihologice pe linie de personal (ex. activități de selecție profesională, evaluarea performanțelor profesionale, utilizarea testelor psihologice etc.), au fost concepute o serie de norme deontologice care să protejeze psihologia ca știință și psihologii, să controleze și să reglementeze legalitatea practicării profesiei de psiholog. (Se știe că practicile diletante, abuzul de testare psihologică sau imixtiunile unor neaveniți în activitățile de diagnoză a resurselor umane ale companiilor sunt, din nefericire, practici curente care aduc numai prejudicii atât profesiei de psiholog cât și psihologiei ca știință). Metodologia psihologică este orientată tot mai mult pe fundamentarea științifică a soluțiilor propuse, pe evitarea unor posibile discriminări de natură etnică, de sex sau vârstă etc.

Automatizarea, înstrăinarea, șomajul, transferările, plafonarea profesională etc. sunt probleme sociale cu care psihologia personalului se confruntă din ce în ce mai des. Pe această linie companiile nu au făcut eforturi prea mari, problema rămânând deschisă. În viitor, ritmul accelerat al schimbărilor sociale va face ca recrutarea și selecția profesională să devină activitatea prioritară de top a organizațiilor. Se estimează că peste puțin timp problemele sociale ale muncii vor produce cele mai multe schimbări la nivelul societății contemporane.

Fiecare concediere, erorile de selecție profesională, eșecurile de repartiție profesională sau instruire, absenteismul și fluctuația personalului sunt fenomene cu profunde implicații economice și care

costă companiile mulți bani. Iată de ce dezvoltarea și implementarea unor strategii de personal fundamentate științific constituie una din preocupările majore ale psihologilor antrenați în problematica deosebit de complexă a psihologiei personalului. Responsabilitatea lor a crescut considerabil în contextul industriei moderne.
Totodată, se poate afirma că psihologia personalului câștigă tot mai mult teren, devenind unul din domeniile de intervenție care interesează companiile din ce in ce mai mult.

Psihologia personalului nu poate fi izolată de politica organizațională, care este și o extensie a politicii statului. Psihologul este un factor de echilibru, un moderator care, integrat corespunzător în viața organizației, va contribui la o mai bună ancorare a acesteia în realitate.

CAPITOLUL 2

CONCEPTE FUNDAMENTALE ÎN MĂSURAREA PSIHOLOGICĂ

Variabile statistice

Analiza statistică a unei populații se face, totdeauna, pornind de la însușirile ei. Orice însușire, proprietate sau trăsătură care poate fi urmărită la toate unitățile statistice din populația investigată se numește *caracteristică statistică*. Valorile pe care le ia o caracteristică statistică se numesc *modalități*.

Exemplul 1. Într-o cercetare privind numărul de ore petrecute săptămânal în fața televizorului, într-o țară, caracteristicile statistice pot fi:

- numărul de ore petrecute la televizor - modalități: numere întregi (numărul de ore se rotunjește la cel mai apropiat întreg);
- sexul - modalități: "masculin", "feminin";
- vârsta - modalități: numere întregi (vârsta în ani împliniți);
- nivelul de școlarizare - modalități: "fără școală", "cu școală primară", "cu școală generală", "cu liceu", "cu facultate";
- mediul de domiciliu - modalități: "urban", "rural".

Trebuie reținut că modalitățile caracteristicii sunt stabilite de cercetător, în conformitate cu felul de manifestare a caracteristicii și cu scopul investigației. Pentru vârstă, de exemplu, modalitățile pot fi:

- vârsta exprimată în ani împliniți

sau

- intervale de vârstă, de exemplu: (0,10], (10,20],..., (90,100].

Sau, pentru mediul de domiciliu, s-ar fi putut considera ca modalități: "municipiu","oraș", "comună", "sat".

Dar, este absolut necesar ca modalitățile unei caracteristici să fie astfel stabilite încât fiecărei unități statistice din populație să i se poată pune în corespondență o modalitate și numai una.

O caracteristică pentru care în populația statistică există o singură modalitate se numește *constantă*. În unele cercetări este important ca o caracteristică să fie constantă și se verifică acest lucru pentru fiecare unitate statistică.

Exemplul 2. Se face un studiu privind influența unei substanțe chimice asupra memoriei. Dar, este posibil ca relația dintre performanțele la memorare și cantitatea de substanță chimică introdusă în organism să fie influențată de sex și de vârstă. Din acest motiv se vor forma eșantioane în care caracteristicile sex și vârstă vor fi constante.

O caracteristică pentru care în populația statistică există cel puțin două modalități se numește *variabilă statistică*. Modalitățile caracteristicii sunt valorile variabilei.

După cum se observă din exemplul 1, unele caracteristici au modalități calitative (exprimate prin cuvinte) - de exemplu sexul, nivelul de școlarizare, mediul de domiciliu -, altele au modalități cantitative (exprimate prin numere) - de exemplu, vârsta.

Descrierea calitativă este rezultatul unei operații de clasificare, adică de repartizare a obiectelor în clase sau categorii, astfel încât obiectele fiecărei clase să fie cât mai asemănătoare între ele, iar obiectele aparținând unor categorii diferite să se deosebească între ele cât mai mult.

Descrierea cantitativă impune efectuarea unei măsurări; aceasta înseamnă atribuirea de numere obiectelor, în conformitate cu anumite reguli, fixate de cercetător. Evident, pe baza măsurilor lor, obiectele pot fi ulterior clasificate.

Pentru ca o variabilă să poată fi utilizată într-o cercetare ea trebuie să fie *definită operațional*, adică trebuie să fie specificate toate activitățile pe care cercetătorul le face ca să o măsoare sau ca să o manipuleze.

Exemplul 3. Definiția operațională a variabilei "consumul de hrană" poate conține următorii pași:

- se înregistrează alimentele consumate în cursul unei zile de o persoană;
- pentru fiecare aliment se caută în tabele valoarea energetică, exprimată în kcalorii/g;
- se înmulțește valoarea energetică a fiecărui aliment cu cantitatea consumată și se adună rezultatele;
- numărul obținut reprezintă măsura "consumului de hrană", exprimată în kcalorii/zi.

Importanța măsurării în psihologie

În școală, în industrie sau în clinică, psihologul este pus în situația să evalueze sau să măsoare diferite calități psihice, cunoștințe, progresul unui demers psihoterapeutic etc. Suntem deci cât se poate de mult familiarizați cu măsurarea, sau mai precis cu diferite tipuri de măsurări, măsurători sau cuantificări care, într-o accepțiune cât se poate de generală, presupun adoptarea unor reguli de măsură, o anumită metrie, criterii de măsură, instrumente cu care se face măsurarea etc. Experiența noastră cotidiană privind măsurarea și diferitele tipuri de măsurători ne spune că este foarte dificil să cuantificăm variabile psihice. Cum putem măsura inteligența, capacitatea creatoare sau temperamentul etc. este o problemă la care

nu se poate răspunde cu ușurință. Un lucru este evident: măsurarea în psihologie diferă mult de aceea pe care o face inginerul când măsoară de pildă înălțimea unui bloc sau lungimea unui pod. Punctul central al psihologiei personalului îl constituie măsurarea diferențelor individuale. Deciziile despre oameni cer aprecierea individualității lor, a cunoștințelor pe care ei le pot câștiga. Aceasta se poate realiza numai prin măsurarea obiectivă și sistematică a patternelor individuale de aptitudini, deprinderi, cunoștințe și interese. O înțelegere adecvată a termenului de "măsurare" este prin urmare crucială pentru interpretarea și evaluarea instrumentelor psihologice, printre care testele psihologice își au un loc bine determinat.

Dacă vizităm un magazin cu produse textile, observăm marea diversitate a costumelor de haine. Ele diferă ca mărime, culoare, fazon etc., drept consecință a deosebirilor existente între cumpărătorii lor, în privința dimensiunilor și a preferințelor. O lege psihologică, pe care fiecare dintre noi o cunoaște, este că oamenii diferă unii de alții. Să ne reamintim un episod din vacanța pe care ne-am petrecut-o pe o plajă de pe Litoralul Mării Negre. Oamenii veniți la odihnă ne oferă o paletă foarte variată de prezențe. Unii afișează un comportament deschis și gregar, alții sunt mai timizi și retrași: câțiva dau dovadă de creativitate în construcția castelelor de nisip, alții sunt mai puțin imaginativi; unii sunt înalți, alții scunzi etc. Ca specialiști în domeniul științelor socio-umane, obiectivul nostru este să surprindem și să descriem această variabilitate, să o înțelegem și să o explicăm, să o prezicem. Prin măsurare reușim să ne apropiem mai mult de obiectivele formulate. O dată ce vom înțelege de ce măsurăm și cum se face măsurarea, tehnicile de măsură vor deveni mult mai ușor de înțeles, vor putea fi utilizate cu mai multă competență.

Ce s-ar întâmpla dacă nu am utiliza măsurarea și instrumente de măsură în psihologie? Aceasta ar însemna să nu putem descrie, compara sau identifica fenomenele contrastante din mediul înconjurător. Nu vom putea să utilizăm nici noțiunile prin care se

încearcă descrierea structurii personalității unui individ (inteligență, caracter, temperament etc.) și nici unitățile de măsură aferente comportamentului uman (coeficient de inteligență, șansă de succes școlar sau profesional, coeficient de învățare etc.). Însăși cercetarea psihologică sistematică nu se va mai putea desfășura. Din fericire, abordările științifice în psihologie s-au impus în timp. Ne rămâne să facem doar unele precizări lămuritoare despre măsurarea psihologică, să vedem ce este aceasta, care-i sunt particularitățile, calitățile și limitele.

Notă:
Unii psihologi vehiculează ideea că statistica -deci, implicit, și măsurarea în psihologie- ar fi desuetă, că simpla intuiție psihologică sau judecata clinică sunt suficiente și edificatoare pentru practica psihologului. Această pretenție este cât se poate de falsă, este o practică diletantă care ascunde mai mult lipsa de competență și abaterea flagrantă de la principiile psihologiei științifice. O discuție interesantă legată de acest subiect, despre noi alternative în psihodiagnoza clinică ne este oferită de Arthur (1969).

Natura măsurării psihologice

Definirea măsurării

Într-un sens cât se poate de general, măsurarea este procesul de atribuire de numere obiectelor sau evenimentelor, în conformitate cu niște reguli bine precizate (Stevens, 1951). Ca un caz particular, măsurarea psihologică este procesul de atribuire de numere persoanelor, astfel încât anumite relații existente între indivizi în privința atributului măsurat să fie reflectate fidel de câteva proprietăți ale numerelor. De exemplu, măsurarea atenției distributive se poate face atribuind fiecărei persoane un număr între 1 și 7, astfel încât ori de câte ori o persoană are o atenție mai bună decât alta să primească un număr mai mare decât aceasta și oricare două persoane care au un același nivel al atenției să primească numere egale.

Să considerăm un concurs de frumusețe. Concurentele trebuie să se încadreze în niște bareme fizice cât se poate de clare. Figura 2-1 ne prezintă "proporțiile perfecte" pentru Miss Univers. După cum se observă, este vorba despre calități fizice (lungimi) pentru evaluarea cărora există instrumente de măsură. Dar, dacă, pe lângă dimensiunile fizice, comisia trebuie să evalueze candidatele și din punctul de vedere al calităților intelectuale, apar o serie de întrebări: Cum să se procedeze pentru a aprecia nivelul intelectual al fiecăreia? Din ce comportamente să se deducă acesta? Ce instrument de măsură să se utilizeze? O rezolvare o poate constitui construirea unei scale de notare de la 1 la 10 pe care fiecare membru al comisiei va nota cât de inteligentă i se pare intervenția fiecărei candidate. În acest fel, calitățile intelectuale ale fiecărei persoane vor fi evaluate prin mai multe note, care s-ar putea să difere între ele. Pe care să o alegem drept măsură a inteligenței, cea mai mare, cea mai mică sau media lor? Altfel spus, ce regulă să formulăm pentru ca, pe baza acestor note, fiecare candidată să primească un număr care să îi cuantifice nivelul intelectual? Indiferent de felul în care s-ar proceda, este clar că atât calitatea, cât și precizia evaluării inteligenței sunt mult mai puțin exacte decât în cazul măsurării calităților fizice. Totuși, prin faptul că atribuie fiecărei persoane câte un număr, conform unor reguli fixate, ambele evaluări satisfac definiția dată de S.S. Stevens măsurării. Menționăm că definiția în cauză nu spune nimic despre *calitatea* măsurării.

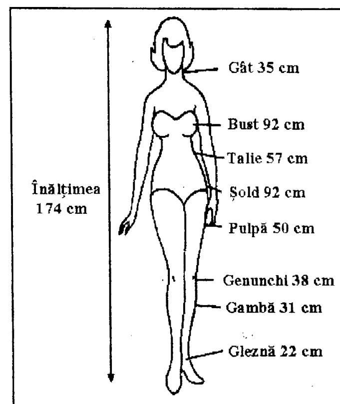
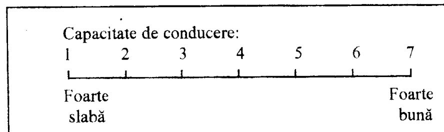


Figura 2-1 "Proporțiile perfecte" pentru Miss Univers.

Reținem că procedurile de măsurare, indiferent că se referă la calități fizice sau psihice, sunt identice. Ele impun să se stabilească în ce constă operația de măsurare, prin ce numere se cuantifică ceea ce se măsoară și care sunt regulile prin care se atribuie fiecărui individ sau obiect un anumit număr. De exemplu, pentru măsurarea anxietății se poate administra fiecărei persoane scala *Anx* din *Inventarul Psihologic California* (operația de măsurare). Aceasta este alcătuită din 22 de itemi dihotomici, deci furnizează un scor cuprins între 0 și

22 (numerele prin care se cuantifică anxietatea). Măsurarea anxietății se face atașând fiecărui subiect scorul realizat la scala *Anx* (regula).

Un exemplu de scală pe care putem face măsurarea unei calități psihice:



Definiția măsurării impune ca prin operația respectivă să se atribuie numere persoanelor, obiectelor sau evenimentelor. Este necesară, însă, următoarea precizare: noi nu măsurăm persoanele, obiectele sau evenimentele, ci caracteristicile lor, caracteristici care, pentru o anumită populație, pot fi constante sau variabile. Astfel, de exemplu, atribuind fiecărui elev al unei clase numărul care reprezintă greutatea sa exprimată în kilograme se realizează o măsurare a elevilor; dar, de fapt, se măsoară caracteristica "greutate" la fiecare persoană.

Măsurarea unei caracteristici se bazează totdeauna pe anumiți *indicatori* ai acesteia. Un indicator este un fapt observabil, care permite să se aprecieze prezența sau absența caracteristicii cercetate, eventual gradul în care această caracteristică este prezentă. Deplasarea acului unui cântar este un indicator al greutății obiectului aflat pe taler. Faptul că pe un teren de fotbal un jucător acumulează un număr mare de penalizări din partea arbitrului este un indicator al comportamentului agresiv al jucătorului în cauză.

Dar, în timp ce caracteristicile fizice și biologice ale obiectelor și indivizilor (ex. sex, culoarea părului) sunt accesibile observației

directe, caracteristicile psihice nu pot fi observate direct. Anxietatea, agresivitatea, creativitatea, inteligența sunt variabile care nu pot fi evaluate direct, ci se *deduc* folosind diverși indicatori ai comportamentelor prin care aceste caracteristici psihice se manifestă. Atribuirea de numere -deci. măsurarea- este posibilă și în acest caz, dar pe o cale indirectă, utilizând o scală de măsură a comportamentelor. Desprinderea proprietăților psihologice din datele observației directe este, probabil, cea mai mare dificultate a măsurării psihologice (Kerlinger, 1986).

Diferențele individuale

În măsurarea psihologică, una din problemele cheie o reprezintă evidențierea diferențelor dintre oamenii. În mod concret, între indivizi există deosebiri reale, relativ stabile, în comportament, interese, preferințe, percepții, convingeri etc.

Unele diferențe sunt așa de mici sau foarte specializate încât au un impact nesemnificativ asupra oamenilor (ex. Vasile colecționează timbre din Grecia, Radu colecționează timbre din Franța). Alte deosebiri individuale, însă, pot afecta succesul profesional sau școlar ori în afaceri (ex. capacitatea de adaptare sau sănătatea psihică etc). Măsurarea sistematică a caracteristicilor psihice la care este importantă evidențierea diferențelor individuale constituie obiectivul major al psihologiei aplicate.

Pentru măsurarea comportamentul uman se proiectează diverse proceduri de traducere a diferențelor individuale în termeni cantitativi. Prin aceste proceduri se alocă numere indivizilor, astfel încât să fie îndeplinite următoarele două condiții :

- ordinea de mărime a numerelor atribuite să corespundă ordinii indivizilor în ceea ce privește comportamentul măsurat;

- diferența dintre numerele atribuite la două persoane oarecare să fie cu atât mai mare, cu cât persoanele respective se deosebesc mai mult între ele.

Testele psihologice sunt exemple de asemenea proceduri. Cu ajutorul lor se pot evidenția caracteristicile psihice în privința cărora doi sau mai mulți indivizi se deosebesc între ei și cele la care nu se constată diferențe importante. De asemenea, testele psihologice pot semnala ceea ce este specific fiecărei persoane, întrucât ele pot indica acele caracteristici psihice la care individul respectiv se deosebește mult de grupul din care face parte (colectivul de muncă sau clasa socială sau grupa de vârstă etc.).

Scale de măsură

Din punct de vedere matematic, *măsurarea* este o funcție care atribuie fiecărui element din mulțimea de persoane, obiecte sau evenimente (mulțimea de definiție a funcției), conform unei anumite reguli, un număr și numai unul (din mulțimea în care funcția ia valori), numit *"măsura elementului"*.

Regulile impuse au rolul de a face ca proprietățile măsurilor elementelor să corespundă proprietăților elementelor. Mai precis, dacă în mulțimea obiectelor este definită o anumită relație (de exemplu, o relație de ordine), atunci trebuie să existe o relație similară în mulțimea măsurilor elementelor, iar dacă două obiecte sunt în relația considerată atunci măsurile lor trebuie să fie în relația similară ei. De exemplu, dacă, din punctul de vedere al caracteristicii cercetate, un obiect A este superior unui obiect B, atunci măsura obiectului A trebuie să fie mai mare decât măsura obiectului B.

Înseamnă, deci, că măsurarea stabilește un izomorfism între mulțimea obiectelor și mulțimea măsurilor obiectelor. În asemenea condiții se pot efectua diverse operații asupra măsurilor obiectelor, iar concluziile desprinse se extind asupra obiectelor.

Fiecare sistem de reguli impus definește câte un tip ori nivel de măsurare sau câte un tip de *scală*. În practica psihologică sunt utilizate patru tipuri de scale de măsură. În ordinea crescătoare a complexității lor, acestea sunt:

- scala nominală;
- scala ordinală;
- scala de interval;
- scala de raport.

Dacă este posibilă definirea unui tip de măsurare, înseamnă că sunt îndeplinite condițiile pentru definirea oricărui alt nivel de măsurare inferior acestuia. De exemplu, dacă s-a putut defini o scală de interval, atunci s-ar fi putut defini și o scală ordinală sau una nominală.

Trebuie reținute, însă, două idei:

- Alegerea unui anumit nivel de măsurare pentru o mulțime de obiecte este dependentă de numărul și tipul relațiilor existente între obiecte și care prezintă interes pentru cercetător. Evident, aceste relații se referă la caracteristica analizată.
- Prelucrările matematice sau statistice care pot fi efectuate asupra măsurilor obiectelor, pentru a cunoaște mai bine obiectele în privința caracteristicii avute în vedere la efectuarea măsurării, sunt dependente de tipul de măsurare utilizat.

Atunci când se iau în considerare mai multe tipuri de relații între obiecte, se poate utiliza un nivel mai înalt de măsurare, care permite efectuarea unor prelucrări mai complexe asupra măsurilor obiectelor și conduce la concluzii mai numeroase și mai precise.

Notăm în continuare cu A mulțimea obiectelor (sau a indivizilor sau a fenomenelor), cu B mulțimea de numere din care se dau valori măsurilor elementelor din A și cu $m:A \rightarrow B$ funcția care realizează măsurarea. Deci, pentru orice obiect $a \in A$, $m(a)$ reprezintă măsura sa.

Scala nominală poate fi utilizată ori de câte ori elementele mulțimii A sunt împărțite în clase sau categorii neordonate, disjuncte două câte două, adică atunci când există o partiție a mulțimii A. Evident, repartizarea în clase are la bază o anumită caracteristică. Mai precis, în fiecare clasă sunt grupate toate obiectele care pot fi considerate echivalente în privința acestei caracteristici.

Pentru oricare două elemente $a,b \in A$ are loc una și numai una dintre următoarele situații:
a și b fac parte din aceeași clasă, adică sunt echivalente (se notează $a=b$)
sau
a și b fac parte din clase diferite, adică nu sunt echivalente (se notează $a \neq b$).

De pildă, categoriile socio-profesionale (atunci când nu se are în vedere o ierarhizare) sau clasele constituite din persoane de același sex fac posibilă utilizarea unei scale nominale.
Deci, pentru definirea unei scale nominale este necesar ca în mulțimea obiectelor, A, să fie definită o relație de echivalență.

Regula impusă la măsurarea nominală este ca toate elementele dintr-o aceeași clasă de echivalență să aibă aceeași măsură și o valoare a măsurii să apară numai la elemente echivalente între ele.

Se observă că o modalitate simplă de definire a unei scale nominale constă în atribuirea de numere distincte claselor de obiecte. Toate obiectele unei clase vor avea aceeași măsură, egală cu numărul atribuit clasei. De exemplu, prin "numărarea" elementelor din A se

realizează o măsurare nominală; în acest caz fiecare clasă este constituită dintr-un singur obiect. Numerele pe care le primesc jucătorii unei echipe de fotbal în cadrul unui joc sunt de asemenea măsuri nominale.

Variabilele care iau valori printr-o măsurare de tip nominal se numesc variabile nominale sau variabile categoriale sau variabile enumerative.

Măsura unui element este o "etichetă" atribuită elementului, care servește doar la denumirea, identificarea sau clasificarea acestuia. Ea este numerică, dar ar fi putut fi exprimată și prin cuvinte. Din acest motiv nu are sens să se efectueze calcule aritmetice cu măsurile nominale.

Prin măsurare nominală se obține o singură informație asupra elementelor mulțimii A: două elemente care au aceeași măsură sunt echivalente între ele. Se pot schimba între ele numerele atribuite claselor fără ca informația pe care ne-o furnizează o măsurare nominală să se schimbe.

Prelucrările statistice obișnuite în cazul variabilelor măsurate printr-o scală nominală constau în calculul frecvențelor absolute sau relative ale claselor. Un singur indice statistic de start poate fi determinat, și anume modul.

În cazul utilizării scalei nominale, datele pot fi reprezentate grafic prin diagrame de structură ori prin diagrame în benzi sau în coloane.

Prelucrările statistice permise pentru datele măsurate prin scale nominale sunt cele bazate doar pe frecvențe: testul χ^2 și unele măsuri ale asocierii, cum sunt coeficienții de contingență.

Scala ordinală impune ca elementele mulțimii A să fie ierarhizate (adică, aranjate în ordine, dar cu permisiunea ca mai multe obiecte să ocupe același loc) în funcție de o anumită caracteristică sau după un anumit criteriu.

Convenim să spunem că un element $a \in A$ este "superior" unui element $b \in A$ dacă în șirul ordonat format elementul a se află înaintea lui b. Notăm acest lucru prin $a>b$. Relația > în cazul unei caracteristici concrete se poate traduce prin "este mai mare decât" sau "este mai bun decât" sau "este preferat lui" etc.

Dacă două elemente, $a,b \in A$, ocupă același loc în șirul ordonat, vom spune că a și b sunt "echivalente" și vom nota $a=b$. Relația = poate însemna, în funcție de context, "este la fel de mare ca" sau "este la fel de bun ca" sau "este apreciat la fel ca" etc.

Pentru a putea defini o scală ordinală, elementele mulțimii A trebuie să îndeplinească următoarele condiții:

C1. Pentru oricare două elemente distincte $a,b \in A$:
-dacă $a=b$ atunci are loc și $b=a$ (simetria relației =);
-dacă $a>b$ atunci nu are loc $b>a$ (antisimetria relației >).

C2. Nici un element $a \in A$ nu poate fi în relația > cu el însuși (ireflexivitatea relației >).
Pentru orice element $a \in A$ este valabilă numai relația $a=a$ (reflexivitatea relației =).

C3. Pentru oricare trei elemente distincte $a,b,c \in A$:
- dacă au loc relațiile $a>b$ și $b>c$ atunci trebuie să aibă loc relația $a>c$ (tranzitivitatea relației >);
- dacă au loc relațiile $a=b$ și $b=c$ atunci trebuie să aibă loc relația $a=c$ (tranzitivitatea relației =).

C4. Pentru oricare două elemente distincte $a,b \in A$ trebuie să aibă loc una și numai una dintre relațiile:
$a>b$, $b>a$, $a=b$

De exemplu, muncitorii unei secții pot fi ierarhizați în funcție de randamentul lor, de la cel mai eficient la cel mai puțin eficient. Vom spune că a este "superior" lui b *($a>b$)* dacă a este mai eficient decât b și vom considera că a și b sunt "echivalenți" *($a=b$)* dacă au același randament. Relațiile > și = definite în acest mod îndeplinesc condițiile *C1, C2, C3* și *C4*, deci permit utilizarea unei scale ordinale.

În schimb, relația definită prin "$a>b$ dacă și numai dacă a îl domină pe b" nu îndeplinește condiția de tranzitivitate : este posibil ca într-o familie mama să îl domine pe tată, tatăl să îl domine pe copil, iar copilul să o domine pe mamă.
Același lucru este valabil și pentru relația definită prin "$a=b$ dacă și numai dacă a și b lucrează bine împreună".

Condiția de tranzitivitate este dificil de îndeplinit atunci când persoanele sunt analizate pe baza unor variabile psihologice cu conținut social (de exemplu, încredere, cooperare, coeziune etc.).

Deci, o scală ordinală poate fi definită atunci când elementele mulțimii A pot fi grupate în clase sau categorii "ordonate". Toate elementele echivalente între ele formează o clasă, iar clasele pot fi aranjate astfel încât toate elementele unei clase sunt în relația > cu toate elementele oricărei clase aflate după ea.
O asemenea partiție a mulțimii A formată din clase "ordonate" se numește *clasament* sau *ierarhie*.

Regulile impuse măsurării ordinale sunt următoarele:
R1. Două elemente, $a,b \in A$, au măsuri egale *($m(a)=m(b)$)* dacă și numai dacă ele sunt echivalente *($a=b$)*.

R2. Pentru oricare două elemente $a,b \in A$ aflate în relația $a>b$ măsura elementului a este mai mare decât măsura elementului b: $m(a)>m(b)$.

O modalitate simplă de definire a unei scale ordinale constă în atribuirea de numere, în ordine descrescătoare a valorilor, claselor ordonate: prima clasă (cea mai bună) primește un număr, a doua un număr mai mic, a treia un număr și mai mic ș.a.m.d. Toate elementele unei clase vor avea aceeași măsură, egală cu numărul atribuit clasei.

Acordând persoanelor note de la 1 la 5, în funcție de performanțele la locul de muncă, cu semnificația 5=foarte bun, 4=bun, 3=mediu, 2=slab, 1=foarte slab, se definește o scală ordinală. Nota primită de fiecare individ este măsura acestuia.

Scalele de atitudini sunt adesea scale ordinale pentru că ele le cer subiecților să noteze nivelul de satisfacție sau de adeziune la o idee prin marcarea diferitelor grade.

Dacă relația $a>b$ indică superioritatea lui a față de b în privința caracteristicii cercetate, atunci măsurarea ordinală ne dă următoarele informații:

- două elemente care au aceeași măsură sunt echivalente (nici unul nu îi este superior celuilalt);
- dintre două elemente cu măsuri neegale, cel care are măsura mai mare este superior celuilalt.

Dacă măsurile obiectelor din mulțimea A sunt transformate printr-o funcție crescătoare, se obține o altă măsurare ordinală a elementelor din A, întrucât noile măsuri păstrează ordinea obiectelor.

Trebuie reținut că măsurarea ordinală nu ne dă informații despre "distanța" existentă între două elemente ale mulțimii A.

Dacă $m(a)-m(b)=m(c)-m(d)$ nu înseamnă că a este tot cu atâta superior lui b cu cât c este superior lui d.

Să considerăm, de exemplu, cazul când elementele lui A sunt elevii unei clase. Aceștia sunt ordonați crescător după cunoștințele lor la matematică și fiecare este măsurat prin locul său. Dacă printre elevi există o persoană cu premii la olimpiade, aceasta va ocupa ultimul loc, deci va avea măsura cea mai mare, dar "distanța" (în ceea ce privește cunoștințele de matematică) dintre ea și penultimul din șir este cu mult mai mare decât "distanța" dintre penultimul și antepenultimul.

Indicele cel mai potrivit pentru descrierea tendinței centrale a datelor măsurate printr-o scală ordinală este mediana. Evident, este posibilă și determinarea modului.

Pentru variabilele măsurate ordinal se poate recurge la calculul rangurilor centile, decile sau cuartile.

Datele măsurate la nivel ordinal se reprezintă grafic prin diagrame în benzi, în coloane sau în batoane.

Pentru verificarea ipotezelor statistice se utilizează teste statistice neparametrice, iar pentru cercetarea relației dintre variabile măsurate prin scale ordinale se calculează coeficienții de corelație a rangurilor: de exemplu, coeficientul *rho* al lui Spearman, în cazul a două variabile, sau coeficienții de concordanță (de exemplu, coeficientul W al lui Kendall), în cazul mai multor variabile.

Scala de interval poate fi utilizată atunci când în mulțimea A se poate stabili un clasament și, în plus, se poate aprecia "distanța" dintre oricare două elemente în privința caracteristicii cercetate. Această "distanță" este o funcție

$$d : AxA \rightarrow R$$

care atribuie fiecărei perechi de elemente din A, (a,b), un număr nenegativ, $d(a,b)\geq 0$, și îndeplinește următoarele trei condiții:

C1. "Distanța" dintre oricare două elemente echivalente este egală cu zero:

$$a=b \Rightarrow d(a,b)=0$$

C2. "Distanța" dintre a și b este egală cu "distanța" dintre b și a:

$$d(a,b)=d(b,a)$$

pentru orice elemente $a,b \in A$ (proprietatea de simetrie).

C3. Pentru oricare trei elemente $a,b,c \in A$, între care există relațiile $(a>b$ sau $a=b)$ și $(b>c$ sau $b=c)$, are loc egalitatea:

$$d(a,c) = d(a,b) + d(b,c)$$

(proprietatea de aditivitate).

Unei scale de interval i se cere ca, pe lângă regulile $R1$ și $R2$ impuse scalei ordinale, să îndeplinească și următoarea regulă:

R3. Pentru oricare patru elemente $a,b,c,e \in A$ are loc
$|m(a)-m(b)| = |m(c)-m(e)|$ dacă și numai dacă $d(a,b)=d(c,e)$

Această regulă înseamnă că scala de interval atribuie unor obiecte egal distanțate între ele în privința caracteristicii cercetate, măsuri echidistante.

Să considerăm un concurs de frumusețe, la care S primește nota 2, M nota 5 și L nota 8. Dacă avem motive să credem că în realitate L este mai frumoasă decât M, care este mai frumoasă decât S, și "distanța" dintre L și M, în privința frumuseții, este egală cu "distanța" dintre M și S, atunci putem afirma că notele acordate constituie o măsurare de tip interval a celor trei candidate.

În cazul scalelor de interval, faptul că un element are măsura egală cu zero nu înseamnă că el este lipsit de caracteristica cercetată. De exemplu, nota zero acordată unei candidate la concursul de frumusețe nu semnifică absența totală a frumuseții.

Exemplul tipic de scală de interval se întâlnește la măsurarea temperaturii. După cum se știe, temperatura este un parametru ce caracterizează starea de încălzire a unui sistem fizic și se determină pe baza variației unor mărimi fizice (volum, rezistență electrică, forță electromotoare etc.), astfel încât unor variații egale ale mărimii fizice le corespund variații egale de temperatură. Termometrul cu care se înregistrează valorile temperaturii este gradat în unități echidistante. Indiferent ce reprezintă distanța dintre două gradații succesive (un grad Celsius sau un grad Fahrenheit sau altceva) diferența de temperatură corespunzătoare distanței dintre două gradații succesive este la fel de mare, fie că este vorba despre gradațiile 2 și 3 sau 10 și 11.

Se obișnuiește să se atribuie unei temperaturi măsura zero, dar aceasta nu înseamnă absența totală a temperaturii. Zero grade Celsius corespunde, de exemplu, la 32 de grade Fahrenheit.

Atât scala Celsius, cât și scala Fahrenheit realizează măsurări de interval.

Din exemplul prezentat se deduce că măsurarea de tip interval fixează arbitrar punctul zero (mărimea caracteristicii căreia să i se atribuie măsura zero) și lungimea unității de măsură ("distanța" dintre două niveluri ale caracteristicii căreia să îi corespundă o diferență a măsurilor egală cu unu).

Dacă măsurile obiectelor sunt transformate printr-o funcție liniară:

$$m_{nou}(a) = p + qm(a), \text{ pentru orice } a \in A$$

unde:
p şi q sunt numere reale, $q>0$,
$m(a)$ este măsura elementului a în scala de interval considerată,
$m_{nou}(a)$ este măsura nouă a elementului a
se obţine de asemenea o măsurare de interval.

Această proprietate a scalelor de interval face posibilă compararea obiectelor pe baza mai multor măsurători de tip interval, chiar dacă fiecare dintre aceste măsurători furnizează valori din alt interval - de exemplu, o măsurare dă valori din intervalul (0,10), iar o alta din intervalul (120,200). Prin transformări liniare se poate ajunge la un interval comun tuturor scalelor.

Datele măsurate printr-o scală de interval se reprezintă grafic prin histograme şi poligoane de frecvenţă. Indicii statistici calculaţi cel mai frecvent sunt media, mediana şi abaterea standard. Se pot utiliza coeficienţii de corelaţie liniară şi se pot aplica testele statistice parametrice (testul t, testul F etc.).

Scala de raport poate fi definită ori de câte ori se poate utiliza o scală de interval şi, în plus, se poate aprecia în mod obiectiv când un element este lipsit de caracteristica cercetată.

De exemplu, putem afirma cu siguranţă că un vehicul care staţionează este lipsit de viteză.

Scalei de raport i se impune ca, pe lângă regulile R1, R2 şi R3 să îndeplinească şi regula:

R4. Un element $a \in A$ are măsura egală cu zero dacă şi numai dacă el este lipsit de caracteristica cercetată.

Atunci când caracteristica măsurată reprezintă spaţiu (lungime, suprafaţă, volum) sau timp (durată de viaţă, timp de reacţie) se poate

stabili în mod obiectiv ce înseamnă "absenţa caracteristicii", deci când măsura să fie egală cu zero. Pentru asemenea variabile se pot defini măsurări de raport.
Deosebirea dintre o măsurare de interval şi una de raport constă în faptul că, la măsurarea de interval, din faptul că $m(a) = 2m(b)$ nu se poate trage concluzia că a este "de două mai bun" (în sensul relaţiei >) decât b, ceea ce este posibil însă la măsurarea de raport. De pildă, indiferent dacă înălţimea persoanelor este măsurată în mm, m sau inch, un copil care are 160 cm este de două ori mai înalt decât unul care are 80 cm. Dacă o maşină se deplasează cu 60 km/oră iar alta cu 30 km/oră, putem spune că prima merge de două ori mai repede decât a doua.

Variabilele ale căror valori se obţin printr-o măsurare de tip interval sau de tip raport se numesc *variabile de tip scor* (Graziano, Raulin, 1993).

Cotele obţinute de indivizi la testele psihologice foarte bine construite, la care o diferenţă de un punct are aceeaşi semnificaţie în privinţa caracteristicii măsurate indiferent că este vorba de diferenţa dintre cota 2 şi 3 sau între cota 31 şi 32, sunt măsurări de interval. Deci variabila măsurată prin cotele testului este de tip scor.

Dar, dacă la un acelaşi test de cunoştinţe, doi indivizi au obţinut 20 şi, respectiv, 40 de puncte nu se poate afirma că al doilea are de două ori mai multe cunoştinţe decât primul. Deci măsurarea cunoştinţelor prin scorurile testului nu este o măsurare de raport.
În general, testele psihologice şi cele de cunoştinţe realizează doar măsurări ordinale ale variabilei la care se referă. Există câteva teste de inteligenţă, bine construite, care realizează măsurări de interval.

Datele măsurate printr-o scală de raport pot fi prelucrate prin orice operaţii aritmetice. Pentru ele sunt permise toate prelucrările statistice amintite la scala de interval. În plus, se poate calcula media

geometrică - aceasta necesită cunoașterea punctului zero real, corespunzător absenței caracteristicii cercetate.

Observație. În unele situații o variabilă poate fi măsurată folosind toate cele patru tipuri de scale. De exemplu, să considerăm mulțimea *A* formată din patru persoane, x, y, z și v, ale căror înălțimi, exprimate în cm, sunt, în ordine: 161, 165, 181, 185.

• Aplicând o măsurare nominală, atribuim fiecărei persoane câte un număr:

x: 1.5, y: 0.6, z: 0.9, v: 6

• Dacă ordonăm persoanele crescător după înălțime și atribuim fiecărei persoane locul pe care îl ocupă, obținem o măsurare ordinală:

x: 1, y: 2, z: 3, v: 4.

• Dacă atribuim fiecărei persoane diferența dintre înălțimea să și 160, realizăm o măsurare de interval:

x: 1, y: 5, z: 21, v: 25

• Dacă fiecărei persoane îi atribuim înălțimea sa, realizăm o măsurare de raport:

x: 161, y: 165, z: 181, v: 185

Tabelul 2-1 ne prezintă o sinteză a diferențelor dintre cele patru tipuri de scale descrise.

Tabelul 2-1
Caracteristicile scalelor de măsură

Scale	Operații	Descriere
Nominală	Egalitate	Categorii mutuale exclusive; obiectele sau evenimentele cad numai într-o clasă; toți membrii aceleiași clase sunt considerați egali; categoriile diferă calitativ, nu cantitativ.
Ordinală	Egalitate Ierarhizare	Ideea de ordonare a intrărilor: obiectul este mai mare sau mai mic ori egal cu altul; orice transformare crescătoare este permisă.
De interval	Egalitate Ierarhizare Unități egale	Aditivitate; se pot fixa distanțe echivalente de-a lungul scalei; este permisă orice transformare liniar crescătoare.
De raport	Egalitate Ierarhizare Unități egale Zero absolut	Este definit un punct de zero absolut; poate fi determinat raportul măsurilor a două obiecte.

Cum considerăm numerele în măsurări

Printr-o operație de măsurare, unei variabile i se atribuie valori numerice. Sunt posibile două situații:

1. Variabila poate lua numai anumite valori numerice. Între oricare două valori posibile există doar un număr fixat de valori posibile. Mulțimea tuturor valorilor posibile este finită sau numărabilă. În acest caz se spune că variabila este *discretă*. Ca exemple, amintim variabilele ale căror valori sunt frecvențe absolute (ex. numărul de studenți dintr-o grupă, numărul de zile ploioase dintr-un an etc.). Atunci când scorul unui test psihologic se obține prin însumarea cotelor itemilor componenți și pentru fiecare item scorul poate lua doar un număr finit de valori (de exemplu, 1 pentru răspuns corect, 0 pentru răspuns omis și -1 pentru răspuns greșit), variabila a cărei valoare pentru fiecare persoană este egală cu scorul obținut de aceasta la test este o variabilă discretă.

2. Variabila poate lua orice valoare numerică dintr-un interval finit sau infinit dat. Între oricare două valori posibile ale variabilei există o infinitate de valori posibile. În acest caz se spune că variabila este *continuă*. În general, toate variabilele care exprimă spațiul (lungime, suprafață, volum), timpul (vârsta, timpul de reacție), masa sau combinații ale acestora (viteza, accelerația etc.) sunt variabile continue.

Dar, în cercetare, se întâmplă adesea ca variabile continue să apară ca fiind discrete. O cauză o constituie imprecizia instrumentelor cu care se face măsurarea (de exemplu, înălțimea persoanelor este înregistrată de obicei în cm). Alteori nu se justifică în cercetare o precizie prea mare (de exemplu, pentru investigarea legăturii dintre orele petrecute la televizor și vârstă este suficient dacă vârsta este exprimată în ani, nu trebuie notată în zile sau ore). În aceste condiții se consideră că fiecare valoare a variabilei reprezintă un interval de valori.

Astfel, dacă valorile posibile sunt

$\{x_1, x_2, \ldots, x_m\}$, $x_2 = x_1 + d$, $x_3 = x_2 + d, \ldots, x_m = x_{m-1} + d$, $d > 0$,

atunci fiecare valoare x_i va fi privită ca un interval de valori, de lungime d, pentru care x_i este:

- centrul:

  $(x_i - d/2, x_i + d/2]$, $i = 1, 2, \ldots, m$ sau $[x_i - d/2, x_i + d/2)$

sau

- extremitatea inferioară:

  $[x_i, x_i + d)$, $i = 1, 2, \ldots, m$;

sau

- extremitatea superioară:

  $(x_i - d, x_i]$, $i = 1, 2, \ldots, m$.

Limitele intervalului construit în acest mod pentru o valoare x_i, se numesc *limite reale ale numărului* x_i.

Nota obișnuită la o lucrare scrisă reprezintă, de obicei, centrul unui interval de valori având lungimea de un punct. Astfel, de exemplu, nota 8 indică faptul că, dacă s-ar fi utilizat pentru cotarea lucrărilor numere zecimale, scorul ar fi fost cuprins între 7.5 și 8.5.

Și vârsta exprimată în ani este interpretată adesea în acest mod. Valoarea 20, de pildă, reprezintă vârsta cuprinsă între 20 de ani împliniți și 21 de ani neîmpliniți.

Alteori, dacă o variabilă discretă poate lua un număr mare de valori se consideră că ea este continuă și se acceptă să fie introdusă în prelucrări statistice specifice variabilelor continue. Astfel de variabile sunt venitul pe familie sau pe întreprindere, beneficiul etc.

Alegerea unei scale de măsură

Cercetarea psihologică necesită, în general, măsurarea variabilelor psihice. Se ridică, deci, întrebarea: Cum să alegem scala de măsură pentru fiecare variabilă? Răspunsul este determinat de două aspecte:

- ce se cunoaște despre relațiile existente între persoane în privința variabilei respective, mai precis ce relații au putut fi definite în mulțimea *A* a persoanelor care vor fi măsurate (de exemplu, dacă persoanele pot fi așezate în ordine ori doar se poate afirmă că două persoane sunt echivalente sau nu);
- cât de multe informații ne sunt necesare despre persoane, ce prelucrări statistice ar trebui efectuate pentru a obține aceste informații.

Se știe că scala nominală ne oferă cea mai redusă cantitate de informație: valorile variabilei reflectă doar deosebiri calitative între persoane. Scala ordinală adaugă o informație suplimentară: ordinea existentă între valorile variabilei este aceeași cu ordinea reală a persoanelor. Scala de interval aduce încă o informație: mărimea distanței dintre două valori ale variabilei corespunde mărimii "distanței" dintre persoane. Scala de rapoarte ne furnizează cea mai mare cantitate de informații: ea ne permite să cunoaștem de câte ori este "mai bun" un individ decât altul.

Dacă relațiile cunoscute între persoanele din mulțimea *A* nu ne permit să alegem scala de măsură care să ne ofere informațiile de care avem nevoie, vom alege acea scală care ne furnizează cea mai mare cantitate de informații.

În funcție de scala pe care o alegem, vom selecționa și procedurile statistice pe care le vom utiliza pentru analiza datelor. Implicit, acestea vor afecta concluziile la care vom ajunge. Cu cât cantitatea de informații prelucrată este mai redusă, cu atât și concluziile vor fi

mai sărace, mai puțin capabile să explice efectele variabilelor independente.

Este logic ca în cercetări să încercăm să folosim cât mai mult posibil scalele de interval sau de raport, deoarece ele măresc cantitatea de informație conținută de date. Totuși, ipotezele pe care le formulăm într-o cercetare limitează posibilitățile de alegere a unei scale sau a alteia.

Măsurarea psihologică

Aptitudinile, atitudinile, interesele, sentimentele, indiferent cât sunt de specifice pentru un individ, reprezintă realități psihologice interne care sunt mascate și inaccesibile prin instrumentele cunoașterii directe. Ele sunt însă obiectivate prin conduite. Firește, putem observa ușor în viața cotidiană cum conduite identice pot avea cauze relativ variate și pot primi sensuri psihologice explicative diferite. Allport (1970) arăta că la nivelul personalității, același fapt de conduită poate fi determinat, la persoane diferite, de proprietăți psihologice diferite. Măsurarea psihologică privește, în linii principale, problema diferențelor individuale cu referire directă la *trăsăturile, particularitățile* sau *dimensiunile* psihologice. Ceea ce numim trăsătură este o etichetă descriptivă aplicată la un grup de comportamente care se află în relație unele cu altele (ex. dominanță, agresivitate, inteligență etc.); ea este mai mult sau mai puțin înnăscută sau dobândită. Bazați pe diferite eșantioane standardizate comportamentale (ex. teste, interviuri etc.), noi deducem de fapt poziția sau locul ocupat de un individ în ce privește dimensiunea de personalitate măsurată. Figura 2-2 ne oferă o imagine a demersului interpretativ sau a principiului metodologic general care stă la baza oricărei cunoașteri și măsurări psihologice.

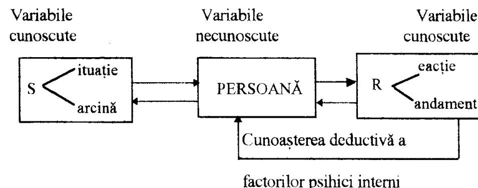


Fig. 2-2 Componentele caracterului deductiv al cunoaşterii psihologice (Kulcsar,1980).

Orice intenţie de cunoaştere sau măsurare a unor parametri psihologici parcurge un drum în care se operează cu o serie de elemente cunoscute şi necunoscute. Iată procedura concretă. Pentru un potenţial candidat la o funcţie managerială psihologul este interesat de cunoaşterea prezenţei sau absenţei calităţilor psihice aferente acestei funcţii. El stabileşte, în consecinţă, o strategie de investigare a subiectului în cauză. Examinarea psihologică debutează prin punerea subiectului într-un context cunoscut (situaţii - probe psihologice sau teste de aptitudini ori personalitate, teste situaţionale etc.). Cu alte cuvinte, psihologul iniţiază o situaţie în care plasează subiectul respectiv si îi impune rezolvarea unei sarcini (parcurgerea unuia sau a mai multor teste de personalitate, de cunoştinţe sau de aptitudini). Este recomandată colectarea a cât de multe date de test sau nontest, adică de comportamente sau răspunsuri (R), în situaţii (S) cât mai variate. Cunoaşterea persoanei este realizată prin examinarea tuturor reacţiilor sau a comportamentelor generate de interacţiunea dintre subiect şi sarcină. Pe bună dreptate sublinia Pavelcu (1974), că "personalitatea nu poate fi înţeleasă decât ca o existenţă în lume, în unitatea organism-mediu, om-lume şi, în acelaşi

timp, ca fiinţă în devenire, într-o neîncetată echilibrare dinamică" (p. 40).

Măsurătorile psihologice, aşa cum apar ele din teste, reprezintă estimări realizate la nivelul unor scale nominale, ordinale sau de intervale. Scalele de rapoarte nu sunt utilizate în măsurările psihologice (acest lucru este totuşi posibil când este vorba de domeniul teoriei răspunsului la item sau al teoriei stărilor şi trăsăturilor latente, care au primit în ultimii ani o extensie considerabilă). Cu alte cuvinte, scorurile obţinute la testele de aptitudini manageriale, inteligenţă sau de personalitate, nu sunt altceva decât estimări pe o scală ordinală. Ele nu ne semnalizează cantitatea de inteligenţă, aptitudini sau trăsături de personalitate ale individului, ci poziţia sau rangul ocupat de un individ în ce priveşte trăsătura măsurată. Pe baza a ceea ce ştim despre scalele utilizate în psihologie, acestea sunt împărţite în intervale egale din raţiuni strict pragmatice (pentru simplul fapt că acest sistem convine scopurilor noastre şi pentru că el, pur şi simplu, funcţionează). Dacă există totuşi dubii serioase cu privire la acest mod de lucru, scorurile brute ale testelor (scorurile derivate direct de la un test, interviu sau scală de evaluare etc.) pot fi transformate prin diferite proceduri statistice în câteva tipuri de alte scoruri care să ne permită utilizarea unor scale uşor interpretabile (Cascio, 1991; Guilford, 1954).

Criterii de evaluare a măsurărilor psihologice

Cronbach şi Gleser (1965) subliniază faptul că obiectivul măsurării psihologice este luarea de decizii. În selecţia profesională, decizia este dacă să admitem sau să respingem un candidat; în repartiţia profesională, dacă să îndrumăm un candidat spre un post de muncă sau spre altul; în diagnoză, care tratament psihoterapeutic să fie adoptat etc.

Măsurările psihologice, arată Cascio (1991), este mai potrivit să fie evaluate în termeni de utilitate socială. Astfel, o problemă importantă

este nu dacă măsurările psihologice folosite într-un context particular sau altul sunt precise sau imprecise, ci cum se poate compara eficienţa lor predictivă cu aceea a altor proceduri şi tehnici existente.

În mod frecvent specialiştii din domeniul psihologiei personalului sunt confruntaţi cu activităţi de selecţie şi cu utilizarea unor proceduri de măsurare psihologică, de interpretare şi de comunicare a intervenţiilor lor. Aceste acţiuni au o importanţă extrem de mare deoarece sunt implicate nemijlocit în orientarea carierei profesionale a unei persoane. Din acest motiv este esenţial ca specialiştii în psihologia personalului să fie bine ancoraţi în intimităţile măsurării aplicative.

Testul: natura sistematică a procedurilor de măsurare în psihologie

Testul psihologic - definiţie. Roşca (1972), defineşte testul ca o probă sau o serie de probe, construite în scopul stabilirii prezenţei (sau absenţei) unui aspect psihic, a particularităţilor de manifestare comportamentală sau a gradului de dezvoltare psihică. Pentru Anastasi (1976), testul reprezintă o măsurare obiectivă şi standardizată a unui eşantion de comportament. În *Grand dictionnaire de la psychologie* (1994) se menţionează că testul psihologic este "o probă, utilizată cu precădere în psihologia diferenţială, care permite descrierea comportamentului unui subiect într-o situaţie definită precis ("instrucţiunile de administrare" ale testului), prin raportare la comportamentul unui grup precis de subiecţi, plasat în aceeaşi situaţie standard. Descrierile se fac de obicei sub formă numerică".

Testul este deci o procedură sistematică utilizată cu scopul comparării comportamentului a două sau mai multor persoane ori a studiului comportamental al unei persoane individuale (Cronbach, 1966). Când afirmăm că procedura de testare psihologică este sistematică sau standardizată, avem în vedere trei aspecte: conţinutul,

administrarea şi cotarea. Referitor la conţinutul unui test, notăm că itemii acestuia trebuie să reflecte şi să măsoare un anumit eşantion comportamental (aptitudinile mecanice, creativitatea etc.). Mai mult, toţi cei care sunt supuşi testării, se vor confrunta cu acelaşi set de itemi sau sarcini. De asemenea, sistemul de administrare a testelor psihologice este şi el standardizat, condiţiile de lucru sau examinare se impune să fie şi ele standardizate, la fel şi durata examinării etc. Procedura de cotare a unui test este, la rândul ei, standardizată prin regulile de notare fixate. Pe scurt, procedurile de operare cu testele psihologice sunt sistematice sau standardizate pentru a diminua distorsiunile posibile şi influenţele nedorite pe care le pot avea unele variabile neesenţiale asupra performanţelor.

O taxonomie posibilă a testelor psihologice. Literatura psihologică despre teste este abundentă. Una din sintezele evaluative cele mai apreciate este pusă la dispoziţia specialiştilor de către Buros (1974). Acesta clasifică testele psihologice existente pe piaţă şi care sunt utilizate în şcoală, industrie şi clinică. Testele sunt evaluate critic după o grilă bine definită de către specialişti în domeniu.

În selectarea unui test pentru a fi utilizat există nişte reguli precise şi o procedură în care psihologul joacă rolul de experimentator şi totodată de filtru. Decizia de selecţie asupra unei probe psihologice se ia în baza unei analize atente a activităţii care urmează să fie investigată. Psihologul trebuie să aibă o idee clară despre ce va fi evaluat, unde şi cum va găsi el ceea ce caută.

Există numeroase sisteme de clasificare a testelor, majoritatea folosind ca punct principal de referinţă definiţia testului. Acest punct de vedere stă şi la baza clasificării pe care o propunem (Figura 2-3).

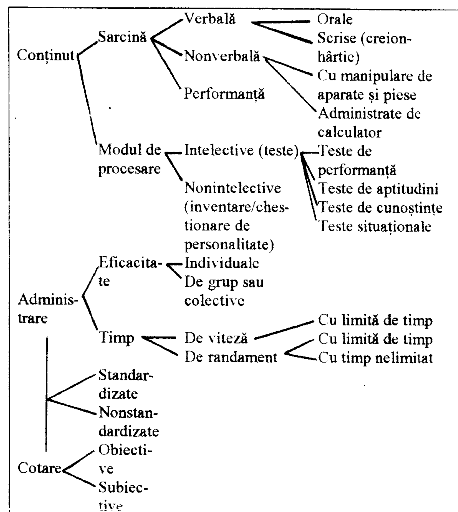


Figura 2-3. Schema de clasificare a testelor psihologice.

Vom examina în continuare, într-o manieră succintă, elementele care dau testului calitatea de probă standardizată: *conținutul, administrarea* și *cotarea*.

Conținutul. Testele pot fi clasificate în primul rând după *sarcina* cu care trebuie să se confrunte persoana examinată. Astfel, unele teste pot fi cu conținut verbal (raționament verbal, de vocabular, completarea unor propoziții) sau cu conținut neverbal (urmărirea mintală a unui labirint, rotirea pe plan mintal a unei figuri). În multe cazuri, subiectului examinat i se poate solicita să opereze la un aparat, să completeze unele figuri lacunare sau să reproducă diferite imagini cu ajutorul unor cuburi etc. Aceste teste sunt cunoscute sub denumirea de probe sau teste de performanță.

Testele se mai pot clasifica și în funcție de modul de *procesare* implicat - aceasta se referă la ce i se cere persoanei examinate să facă pentru a da un răspuns. Testele intelective sunt destinate măsurării calităților intelectuale, fiind frecvent dihotomizate în teste de performanță (achievement) și de aptitudini. În general, ele solicită subiectului să fie performant pe o anumită sarcină sau să ofere o informație concretă într-un anumit domeniu. Atât testele de aptitudini, cât și cele de performanță identifică o abilitate sau capacitate, ori calitate (aptitudini matematice, mecanice sau deprinderi de calcul, cunoștințe de geografie). Ele se pot însă deosebi dacă le privim prin prisma orientării lor. Astfel, *testele de performanță* sunt orientate pe un comportament sau pe o activitate care a avut loc în trecut (învățarea unei limbi străine, deprinderi de depanare auto etc.). *Testele de aptitudini* sunt orientate pe predicția unui comportament apropiat sau mai îndepărtat - ce este individul capabil să învețe pe baza unui program de instruire (ex. aptitudini de operare la calculator, aptitudini muzicale etc.). Desigur, aceste două tipuri de măsurători nu se exclud reciproc; ceea ce un individ a realizat în trecut este frecvent un bun indicator a ceea ce se așteaptă

de la el să facă în viitor. *Testele de cunoştinţe* măsoară nivelul pe care o persoană l-a atins în urma unui proces de învăţare sau cât de multe cunoştinţe a achiziţionat un individ într-un timp dat (în general astfel de teste au fost dezvoltate în şcoală, dar tot atât de bine ele se utilizează în industrie, armată etc.) Distincţia între testele de aptitudini şi cele de cunoştinţe este mai mult una de nuanţă decât de conţinut. De fapt, orice test psihologic poate fi un test de aptitudini dacă el este capabil să prognozeze un comportament viitor. De asemenea, orice test poate fi considerat un test de cunoştinţe dacă poate reflecta cât de mult va putea învăţa sau a învăţat subiectul. Distincţia între testele de aptitudini şi cunoştinţe este determinată, în fond, de sensul utilizării lor. *Probele de lucru* şi *testele* sau *exercitiile situaţionale* constituie o altă categorie de teste intelective. Probele de lucru sunt o replică în miniatură a unei activităţi de muncă. Ele sunt utilizate mai ales în psihologia muncii şi militară ca instrumente de selecţie profesională. Numeroase studii au ajuns la concluzia că acest tip de teste are o valoare predictivă mai mare decât cele clasice. Din acest tip de teste fac parte şi probele *in-basket* şi *jocurile economice* destinate formării profesionale.

O clasă aparte de teste sunt cele pe care le-am grupat în categoria probelor nonintelective. Aceste teste sunt destinate măsurării diferitelor aspecte ale personalităţii (interese, valori, atitudini, trăsături temperamentale etc.). În general, ele se bazează pe expunerea unor convingeri sau atitudini ("Eu cred că ..."; "Un lider ar trebui să ..."; Mă simt ..."). În mod obişnuit, aceste relatări (self-report) sunt grupate sub denumirea de *inventare* în timp ce probele de măsurare a aptitudinilor şi cele de performanţă sunt cunoscute ca *Teste*.

Administrarea. După modul în care sunt administrate testele, ele pot fi clasificate în termeni de eficacitate a examenului psihologic şi în funcţie de unele exigenţe legate de timpul de examinare. Examenul psihologic individual este considerat neeconomic, subiecţii fiind

testaţi unul câte unul, procedura consumând timp, fiind, în consecinţă, neeconomică. Testarea colectivă presupune examinarea simultană a mai multor subiecţi. Ambele moduri de examinare prezintă o serie de calităţi şi limite. Examenul psihologic individual permite psihologului să stabilească un contact mai bun cu subiectul, cooperarea cu acesta se realizează mai rapid, interesul sau motivarea subiectului pentru examinare este superioară. Totuşi, decizia psihologului în contextul examenului psihologic individual este mai încărcată de subiectivitate, examenul psihologic în sine este calificat ca fiind mai puţin standardizat. Condiţiile de examinare psihologică colectivă posedă o notă de standardizare superioară. Dar, unele situaţii particulare ca oboseala, anxietatea, starea momentană a sănătăţii etc., sunt mai greu de detectat în condiţiile examenului colectiv. Testarea individuală se potriveşte mai bine psihologului clinician şi de consiliere în anumite activităţi de orientare şcolară şi profesională; testarea colectivă este preferată în selecţia profesională când se operează cu grupuri mari de subiecţi, mai ales în examinările de tip *screening*. Screening-ul se referă la o categorie de teste sau proceduri simple de identificare rapidă a unor persoane care posedă anumite aptitudini. Obişnuit, astfel de teste sunt administrate înainte de a se lua o decizie de includere într-o acţiune de testare psihologică propriu-zisă. Testele de *screening* lucrează cu o marjă largă de eroare, rolul lor fiind acela de filtrare grosieră a subiecţilor.

În construcţia testelor, ca şi în interpretarea performanţelor acestora, timpul intervine ca o condiţie importantă. Testele de *viteză* (ex. barare de numere sau figuri, identificarea unor figuri identice etc.) conţin, în general, un grup de itemi uşor de rezolvat, timpul de rezolvare fiind însă limitat, astfel încât nici un subiect să nu poată rezolva integral testul. Testele de *randament* se caracterizează prin aceea că timpul de rezolvare este suficient de generos pentru a permite parcurgerea integrală a testului de către subiect. Uneori, administrarea acestui tip de teste nu este condiţionată de timp. La astfel de teste itemii nu au o pondere egală, ei urmează adesea o linie

crescătoare în dificultate; nimeni nu poate obține un punctaj egal cu acela al numărului de itemi a testului (Anastasi, 1982). Atât testele de viteză, cât și cele de randament sunt astfel construite încât să nu permită realizarea unui scor maxim (egal cu numărul itemilor testului). Acest lucru este necesar în vederea obținerii unei imagini mai clare despre persoana testată, a nivelului aptitudinal sau de abilitate atins într-o activitate sau alta. Totodată, este facilitată posibilitatea unei ierarhizări cu mai multă acuratețe a subiecților. În practică, arată Cascio (1990), distincția între teste de viteză și de randament este numai una de grad, cele mai multe teste depinzând de ambii factori în proporții variabile.

Teste standardizate și nestandardizate. Testele standardizate se caracterizează prin existența unor instrucțiuni precise de administrare și cotare. Aceste caracteristici sunt importante în activitatea de comparare a diferitelor persoane între ele. Procedura de standardizare presupune administrarea testului unui eșantion reprezentativ, obișnuit la câteva sute de subiecți, similari celor pentru care testul a fost proiectat (ex. elevi, adulți, personal industrial etc.). Pe acest eșantion standardizat se stabilesc *norme* sau etaloane care reprezintă cadrul de referință pentru interpretarea performanțelor la test. Etalonul testului ne indică nu numai media și dispersia performanțelor eșantionului de subiecți, ci, adesea, și maniera de repartiție a cotelor sau scorurilor testului de-a lungul intervalului de valori ocupat (este vorba despre pozițiile și lungimile claselor centile, decile sau a claselor normalizate).

Testele nestandardizate privesc mai ales categoria probelor de cunoștințe construite de către cadrele didactice într-o manieră neformală și destinate, obișnuit, unei singure administrări (extemporale sau lucrări de control, teze).

Cotarea. Tehnica de cotare a unui test poate fi obiectivă sau subiectivă. În primul caz, performanța la test se apreciază după o procedură fixă, impersonală, de către un simplu operator sau cu ajutorul unei mașini de cotare ori prin scanare și cotare automată prin intermediul unui *software* dedicat, pe calculator. Testele și unele chestionare de personalitate, de genul celor proiective (completare de fraze, pete de cerneală etc.), prezintă în cotare o doză mare de subiectivitate. Ele introduc ceea ce se numește "varianța persoanei care cotează proba". Acesta este motivul pentru care astfel de teste sunt utilizate cu precădere de către psihologii clinicieni sau de unele cabinete de consiliere și mai puțin în mediul industrial.

Precizări

În selecția unui test sau a altui instrument psihodiagnostic, mai pot fi luate în considerare și aspecte ca denumirea, dimensiunea aptitudinală sau de personalitate măsurată, costul, demersul interpretativ sau validitatea de aspect. Primul pas în procedura de selecție a unui test pentru a fi utilizat în mediul industrial este analiza muncii prin care ni se spune ce aptitudini sau abilități trebuie măsurate. Dar, niciodată un instrument psihodiagnostic nu trebuie ales la întâmplare, pe motivul simplu că îl avem și știm să operăm cu el. În măsură egală, alegerea unui test depinde și de obiectivul examinării - selecția de personal, repartiția profesională, expertiza post accident, consiliere sau cercetare.

Formularea clară a obiectivului examenului psihologic ne va concentra atenția asupra găsirii celei mai adecvate strategii de măsurare legat de forma de prezentare, tipul și varietatea scorurilor de care avem nevoie. De pildă, o companie dorește angajarea unui anumit număr de electricieni. Candidații depășesc cu mult locurile disponibile. În acest caz, activitatea desfășurată este una de selecție profesională, dar strategia de lucru este diferită în funcție de obiectivele urmărite. Într-o astfel de situație este indicat să apelăm

inițial la un screening care să realizeze o primă filtrare a candidaților. Examinarea se poate derula în două etape distincte: examinarea îndeplinirii criteriilor de angajare pe CV-urile trimise (criteriile pot fi legate de experiență, vârstă, sex ș.a.). Cei admiși pot fi supuși în continuare unui al doilea filtru, constituit dintr-un test de aptitudini generale (ex. *Testul de personal Wonderlic*) care să permită ierarhizarea subiecților pe baza unui scor global. În fine, o dată lotul de electricieni admis, ei vor fi repartizați la diferite locuri de muncă. Pentru aceasta se va utiliza un test de cunoștințe dublat de unul de aptitudini speciale care să investigheze adecvarea pentru diferite domenii de intervenție pe linie electrotehnică, specific companiei în cauză.

Frecvent, costul unei proceduri de selecție sau de examinare psihologică periodică este un factor important pentru multe companii. Există proceduri de testare de mare precizie, dar foarte costisitoare și altele mai ieftine, dar de o precizie mai mică. Un examen individual este mult mai costisitor decât unul colectiv. Din acest motiv, proiectarea unei examinări psihologice trebuie să includă o analiză logistică de oportunitate în care vor intra: costul materialelor de testare, costurile aferente spațiului de examinare, energia consumată, timpul afectat examinării, personalul de specialitate și ajutător implicat în examinare, publicitatea etc.

Managerii comit frecvent erori când doresc să apeleze la serviciile psihologilor și ale psihologiei. Ei consideră că dacă un test poate fi administrat de o persoană cu o pregătire minimă, și interpretarea acestuia este accesibilă oricui. De fapt, interpretarea testului include mai mult decât o simplă relatare scrisă sau verbală a scorului obținut de o persoană sau un grup de indivizi care au fost supuși examenului psihologic. Interpretare performanțelor sau a rezultatelor unui examen psihologic cere cunoștințe ample care se referă la cunoașterea valorii și a limitelor probelor utilizate (fidelitatea, validitatea etc.), la structura lotului examinat, contextul situațional în care a avut loc

examinarea, specificul strategiei de examinare etc. Necunoașterea problemelor amintite conduce la interpretări eronate, la comiterea de greșeli și de aici la nemulțumirea beneficiarilor și la discreditarea aplicațiilor psihologice.

Validitatea de aspect

Orice test psihologic sau educațional care este destinat utilizării în practică trebuie să fie perceput de cei examinați cu el ca fiind adecvat domeniului în care este aplicat. Această calitate a testelor este denumită **validitate de aspect** (C. I. Mosier, după Nevo, 1993).

De pildă, pe baza rezolvării de scheme logice lacunare se pot construi atât teste de aptitudini de programare, cât și teste de aptitudini pentru electronică. Pentru ca cele două tipuri de teste să aibă validitate de aspect, itemii din care ele sunt alcătuite vor fi diferiți, fiind specifici fiecărui domeniu profesional.

Validitatea de aspect nu este inclusă printre celelalte tipuri de validitate (Standardele APA din 1974) și nici nu le poate înlocui.
O definiție operațională a validării de aspect a fost prezentată de L. Guttman și S. Shye (Nevo, 1993) astfel: "Se face o măsurare a validității de aspect atunci când o persoană testată /un utilizator neprofesionist /un individ interesat, utilizând o tehnică absolută /relativă apreciază un item / un test / o baterie de teste drept foarte potrivit /.../ nepotrivit pentru utilizarea care îl interesează".

Trebuie observat că aprecierea validității de aspect nu este realizată de experți și pentru măsurarea sa nu se folosesc metode psihometrice. Dacă aprecierea validității de aspect este făcută de persoanele testate, ele trebuie să își spună părerea imediat după terminarea testului, când impresiile sunt încă proaspete.

Există două tipuri de tehnici pentru evaluarea validității de aspect:

- tehnica absolută: cel care apreciază cât de potrivit este testul pentru scopul propus folosește o scală cu 5 puncte (5=foarte potrivit, 4=potrivit, 3=mediu, 2=nepotrivit, 1=foarte nepotrivit);
- tehnica relativă: aprecierea testului se realizează prin comparare cu alte teste.

Testele care au o validitate de aspect bună au șanse mari să le inducă subiecților o motivare pozitivă înainte și în timpul administrării testului, să reducă insatisfacția și impresia de injustiție în fața scorurilor mici și să convingă persoanele care utilizează teste psihologice sau educaționale să le implementeze și pe acestea.

CAPITOLUL 3

FIDELITATEA ȘI VALIDITATEA TESTELOR

Consistență și predicție

Întotdeauna când alegem o procedură de măsurare a unei calități psihice sau a mai multora, pretindem ca instrumentul de măsură utilizat să posede o serie de calități care să-i confere o mai mare precizie în măsurare, încredere și obiectivitate în rezultatele măsurării. Obișnuit, în astfel de cazuri, se fac referiri la două caracteristici: ***fidelitatea*** și ***validitatea*** mijloacelor psihodiagnostice.

Orice activitate de examinare psihologică sau de măsurare cu caracter psihologic are ca obiectiv principal luarea unor decizii cu privire la persoanele examinate. În acest context, rezultatele măsurătorilor trebuie să ofere soluții interpretabile, să fie utile. Să examinăm o situație tipică de selecție. O școală dorește alcătuirea unei echipe de atletism, alergători la 100 m garduri. Elevii dornici să facă parte din această echipă care să reprezinte școala într-o competiție zonală trebuie să treacă niște teste pentru a face dovada aptitudinilor solicitate de proba sportivă în cauză. Printre alte verificări, una presupune încadrarea într-un barem de maximum 14 secunde de parcurgere a celor 100 m garduri. Cu alte cuvinte, orice concurent care realizează un timp de alergare de până la 14 secunde va putea intra în componența echipei. În acest sens, pentru fiecare concurent se ia ca bază media timpilor realizați în trei curse (în unele situații similare se ia ca bază timpul cel mai bun). Pentru evaluator este important să intre în posesia unei imagini cât mai "reale" despre aptitudinile implicate în proba amintită. Potențial, vom găsi în fiecare strategie de selecție multe surse de eroare care pot distorsiona "adevărata" realitate. Comportamentul uman fluctuează de la un moment la altul și de la o situație la alta. În plus, procedura de

măsurare în sine conţine ea însăşi numai un eşantion din posibilele situaţii cu care se vor confrunta atleţii în condiţii reale de concurs (indispoziţie, timp nefavorabil, influenţa altitudinii etc.).

Psihologul, în activitatea pe care o prestează, este preocupat de reducerea surselor de eroare care pot influenţa într-un fel sau altul măsurarea. Erorile de măsurare au cel puţin două consecinţe majore asupra unui examen psihologic: scad eficienţa intervenţiei psihologice (avem în vedere aspectele calitative, obiectivitatea rezultatelor, precizia cu care este măsurat ceea ce se doreşte să se măsoare) şi alterează cadrul etic în care se înscrie intervenţia respectivă (se înţelege prin aceasta credibilitatea, imparţialitatea măsurărilor şi a deciziilor). Procedura de măsurare utilizată trebuie să fie una de încredere, consistentă şi stabilă în timp, cu alte cuvinte, cât mai fidelă posibil.

FIDELITATEA

În vorbirea curentă, termenul de fidelitate este folosit pentru a descrie persoane sau lucruri. O persoană fidelă este aceea pe care se poate conta (de exemplu, un angajat fidel îşi termină lucrul său la timp şi execută sarcinile care i se dau la nivelul maxim al aptitudinilor sale). Cuvântul fidelitate este folosit şi pentru a descrie o calitate particulară a numerelor obţinute măsurând caracteristici ale persoanelor sau lucrurilor. În această utilizare, fidelitatea reprezintă consistenţa, consecvenţa măsurătorilor realizate.

În teoria testelor, conceptul de fidelitate a cunoscut diverse accepţiuni pe parcursul timpului. Astfel, de exemplu, în 1954 şi în 1966, Standardele APA identificau fidelitatea cu consistenţa şi stabilitatea ("Fidelitatea se referă la precizia -consistenţa şi stabilitatea- măsurării realizate de test"), în timp ce în 1974, fidelitatea este privită ca şi

"gradul în care rezultatele testării pot fi atribuite surselor sistematice de varianţă" (APA Standards, 1974). Termenii de "consistenţă" şi "stabilitate" continuă să fie utilizaţi, dar nu în sensul de consistenţă internă şi de stabilitate temporară a unui anumit comportament, ci mai degrabă ca reproductibilitate a rezultatelor cercetării ştiinţifice. ***Fidelitatea*** este *"gradul în care scorurile testului sunt consistente sau repetabile, adică gradul în care ele nu sunt afectate de erorile de măsură"* (APA Standards, 1985).

Nici una din măsurările efectuate în sfera factorului uman, indiferent dacă priveşte calităţi fizice sau psihice, nu este complet consistentă. Dacă repetăm o măsurare la una şi aceeaşi persoană în două, trei sau mai multe momente, rezultatele obţinute vor fi mai mult sau mai puţin diferite. De exemplu, dacă ne vom cântări dimineaţa, la prânz şi seara, sau toamna, iarna şi primăvara, vom obţine valori apropiate, dar totuşi deosebite. Dacă examinăm o persoană cu testul *Domino 48* de două ori, la interval de câteva săptămâni, este posibil să se obţină un CI de 115 cu prima administrare şi unul de 118 la a doua. Deoarece scorurile realizate de o aceeaşi persoană nu sunt identice în cele două ocazii, se spune că ele sunt insuficient de consistente sau de robuste. Tehnicile prin care se studiază consistenţa măsurătorilor făcute prin teste, indiferent că este vorba despre teste de aptitudini, inventare de personalitate sau teste de cunoştinţe, sunt cunoscute sub denumirea de tehnici de estimare a fidelităţii.

Fidelitatea testelor psihologice are repercusiuni nemijlocite asupra deciziilor pe care psihologul le ia cu privire la oameni. Să ne imaginăm ce se poate întâmpla dacă într-o acţiune de selecţie este folosit un test de reprezentări spaţiale lipsit de consistenţă, în sensul că aceiaşi candidaţi care au realizat scoruri mici cu o săptămână în urmă, în prezent obţin scoruri mari, în schimb realizează scoruri mici cei care iniţial obţinuseră scoruri mari.

Definiția fidelității în termeni matematici

Se consideră două mulțimi :

P = o mulțime de persoane;

M = o mulțime de măsurători (teste sau itemi de test).

Ambele mulțimi pot fi infinite - dar numărabile- sau pot conține un număr finit de elemente; în particular, ele pot fi formate dintr-un singur element.

Rezultatul aplicării unei măsurători $m \in \mathbf{M}$ asupra unei persoane $p \in \mathbf{P}$ poate fi exprimat sub forma unui număr real, care se notează cu x_{mp}, și se numește *scor observat*. Valoarea lui depinde de condițiile particulare în care se face măsurarea. x_{mp} este doar una dintre valorile care s-ar putea obține pentru măsurarea $m \in \mathbf{M}$ și persoana $p \in \mathbf{P}$. Raționând în acest mod, putem interpreta valoarea x_{mp} ca fiind o realizare a unei variabile aleatoare pe care o notăm cu X_{mp}. Repartiția variabilei aleatoare X_{mp}, dată de toate scorurile observate care s-ar putea obține prin aplicarea, în diverse condiții, a măsurătorii m asupra persoanei p, cu probabilitățile corespunzătoare, nu este cunoscută. În primul rând, este imposibil chiar și numai de imaginat, cu atât mai mult de realizat practic, toate situațiile particulare în care ar putea fi efectuată măsurarea. Pe de altă parte, două scoruri observate nu sunt întotdeauna rezultatul unor experimente independente. O asemenea situație poate să apară, de exemplu, atunci când răspunsurile date la o administrare a unui test sunt memorate și reproduse mecanic la administrările ulterioare.

Pentru variabila aleatoare X_{mp} vom nota valoarea medie cu t_{mp}:

$$t_{mp} = \mathbf{E}(X_{mp})$$

Întrucât repartiția variabilei aleatoare X_{mp} nu este cunoscută, valoarea medie t_{mp} nu poate fi calculată în practică. Dar, teoretic, ea există. Vom presupune că ea este finită, adică $t_{mp}<\infty$. Acest lucru nu contravine realității, căci scorurile observate, pe baza cărora se calculează t_{mp}, fiind rezultate ale unor măsurători psihologice, iau valori într-un interval finit al axei reale.

Pentru o măsurătoare m și o persoană p numărul real t_{mp} se numește *scor real*. El poate fi interpretat ca valoare unică a unei variabile aleatoare constante T_{mp}. Guilford (1965) apreciază că scorul real este valoarea care s-ar obține dacă măsurarea ar fi făcută cu un instrument perfect, aplicat în condiții ideale.

Diferența dintre un scor observat, x_{mp} și scorul real t_{mp} se numește în mod obișnuit *eroare de măsură*:

$$e_{mp} = x_{mp} - t_{mp}$$

e_{mp} poate fi considerat drept o valoare a variabilei aleatoare E_{mp} definită prin relația :

$$E_{mp} = X_{mp} - T_{mp}$$

Nici pentru această variabilă aleatoare nu poate fi cunoscută repartiția. Dar, din definiția scorului real și a erorii de măsură rezultă că :

$$\mathbf{E}(E_{mp}) = 0$$

adică, pentru o persoană p și o măsurătoare m, media erorilor de măsură este egală cu zero.

Analizând relația existentă între variabilele aleatoare X_{mp}, T_{mp} și E_{mp} :

$$X_{mp} = T_{mp} + E_{mp}$$

se poate deduce faptul că scorurile observate la un test (x_{mp}) reflectă influența a două feluri de factori :

- factori care contribuie la *consistență* (care furnizează valoarea t_{mp}) : caracteristicile stabile ale individului sau ale situației în care se face măsurarea;
- factori care contribuie la *inconsistență* (care produc valorile e_{mp}): caracteristici ale individului sau situația de examinare care afectează scorurile observate, dar care nu au nimic comun cu ceea ce se încearcă să se măsoare.

Variabilele aleatoare X_{mp}, T_{mp} și E_{mp} au fost construite pentru o anumită măsurătoare $m \in \mathbf{M}$ și o persoană fixată $p \in \mathbf{P}$. Să aplicăm acum măsurătoarea m, în condiții standardizate, tuturor persoanelor

din **P**. Să notăm cu X_{m^*} variabila aleatoare a cărei repartiţie este dată de toate scorurile observate care s-ar putea obţine în populaţia **P**, cu probabilităţile corespunzătoare.

Variabila aleatoare X_{m^*} este compusă, de fapt, din două variabile aleatoare independente, corespunzătoare următoarelor două ipoteze:

- membrii populaţiei **P** au fost extraşi la întâmplare dintr-o populaţie mai mare;
- măsurătoarea făcută fiecărui individ furnizează o observaţie a varia-bilei aleatoare specifice persoanei, X_{mp}.

Vom nota cu T_{m^*} variabila aleatoare care are ca valori scorurile reale ale persoanelor $p \in$ **P** şi cu E_{m^*} variabila aleatoare corespunzătoare erorilor de măsură care se pot obţine în populaţia **P**.

În populaţia **P** sau în orice subpopulaţie a lui **P** are loc relaţia (Lord & Novick, 1968):

$$X_{m^*} = T_{m^*} + E_{m^*}$$

Din această formulă se deduce că, în întreaga populaţie **P**, deosebirile dintre scorurile observate (valorile variabilei X_{m^*}) provin din două surse:

- deosebirile existente între persoane în privinţa caracteristicii măsurate de test (valorile variabilei T_{m^*});
- erorile de măsură (valorile variabilei E_{m^*}).

În studiul fidelităţii se presupune că erorile de măsură sunt aleatoare şi nu sistematice. Deosebirea dintre cele două tipuri de erori se poate înţelege din următorul exemplu:

Se măsoară în mod repetat greutatea unei persoane. Dacă la fiecare cântărire persoana se urcă pe cântar având în mână o carte de un kilogram, greutatea sa va fi înregistrată greşit. Valorile observate vor fi afectate de o eroare care se produce în acelaşi mod la fiecare măsurare. O asemenea eroare se spune că este sistematică.

Erorile sistematice au efect doar asupra valorilor scorurilor observate, nu şi asupra valorilor scorurilor reale sau asupra a ceea ce se numeşte eroare de măsură. Erorile de măsură sunt complet nepredictibile, indiferent de cunoştinţele pe care le avem despre persoanele măsurate sau despre procesul de măsurare.

Pătratul coeficientului de corelaţie liniară între scorurile reale şi scorurile observate ale testului *m* în populaţia **P**, $r^2(X_{m^*},T_{m^*})$, se numeşte *coeficient de fidelitate al testului m* (Lord & Novick, 1968). Fiind pătratul unui coeficient de corelaţie, coeficientul de fidelitate este cuprins întotdeauna între 0 şi 1.
Din relaţia (Lord & Novick, 1968):

$$r^2(X_{m^*},T_{m^*}) = \sigma^2(T_{m^*}) / \sigma^2(X_{m^*}) = 1 - \sigma^2(E_{m^*}) / \sigma^2(X_{m^*})$$

rezultă :

a. Coeficientul de fidelitate este egal cu 1 dacă şi numai dacă $\sigma^2(E_{m^*})=0$. Deoarece are loc egalitatea $\mathbf{E}(E_{m^*})=0$ (Lord & Novick, 1968), varianţa erorilor de măsură poate fi egală cu zero numai dacă toate erorile de măsură sunt nule. Deci, un test are coeficientul de fidelitate egal cu unu dacă şi numai dacă scorurile observate coincid cu scorurile reale, adică dacă testul este o măsură foarte precisă.

b. Coeficientul de fidelitate este egal cu zero dacă şi numai dacă $\sigma^2(T_{m^*})=0$, deci toate scorurile reale sunt egale.

Este evident că pentru a putea defini fidelitatea unui test trebuie ca $\sigma^2(X_{m^*})>0$, adică scorurile observate la test să nu fie toate egale.

O formulă aplicată frecvent în practică pentru calculului coeficientului de fidelitate necesită administrarea, pe lângă testul *m*, a unui test paralel cu *m*.

Două măsurători distincte $a \in \mathbf{M}$ și $b \in \mathbf{M}$ se numesc *măsurători paralele* dacă pentru orice persoană $p \in \mathbf{P}$:

a. scorul real este același :

$$t_{ap} = t_{bp}$$

b. erorile de măsură au aceeași abatere standard :

$$\sigma(E_{ap}) = \sigma(E_{bp})$$

(Lord & Novick, 1968).

Deci două măsurători paralele măsoară exact același lucru, în aceeași scară ($t_{ap}=t_{bp}$) și îl măsoară la fel de bine ($\sigma(E_{ap}) = \sigma(E_{bp})$)

Întrucât repartițiile variabilelor aleatoare X_{ap} și X_{bp} pentru persoanele p din **P** nu sunt cunoscute, paralelismul măsurătorilor a și b nu poate fi verificat utilizând definiția. Din acest motiv, în practică, studierea paralelismului a două teste se face prin verificarea, pe baza scorurilor observate în eșantionul căruia i s-au administrat testele, a următoarelor egalități :

(1) $\mathbf{E}(X_{a^*}) = \mathbf{E}(X_{b^*})$

(2) $\sigma^2(X_{a^*}) = \sigma^2(X_{b^*})$

(scorurile observate în populație au aceeași medie și aceeași dispersie)

(3) $r(X_{a^*},X_{c^*}) = r(X_{b^*},X_{c^*})$ pentru orice măsurătoare $c \in \mathbf{M}$.

(cele două teste corelează la fel cu orice alt test)

Trebuie menționat, însă, că egalitățile de mai sus reprezintă numai condiții necesare, nu și suficiente pentru paralelism. Dacă ele nu sunt verificate, atunci cu certitudine măsurătorile nu sunt paralele. Dar, realizarea acestor egalități pentru eșantionul testat nu asigură paralelismul.

Dacă $a \in \mathbf{M}$ este o măsurătoare paralelă cu m și dacă pentru orice persoană $p \in \mathbf{P}$ scorurile observate X_{ap} și X_{mp} (sau, echivalent, erorile

de măsură E_{ap} și E_{mp}) sunt variabile aleatoare independente, atunci fidelitatea poate fi calculată cu formula:

$$r^2(X_{m^*},T_{m^*}) = r(X_{a^*},X_{m^*})$$

De aici rezultă că un test m poate avea coeficientul de fidelitate egal cu unu dacă și numai dacă $X_{m^*}=X_{a^*}$. Putem spune, deci, că un test $m \in \mathbf{M}$ este fidel atunci când :

- este precis (lipsit de erori)

sau, echivalent,

- orice persoană obține scoruri egale la testul m și la o formă paralelă a acestuia, a, cele două teste fiind administrate astfel încât erorile de măsură E_{mp} și E_{ap} pentru orice persoană $p \in \mathbf{P}$, să fie independente.

Coeficientul de fidelitate servește la aprecierea preciziei unui test, dar nu este suficient pentru descrierea acestuia ca instrument de măsură. O anumită valoare a coeficientului de fidelitate poate fi obținută pentru variante diferite ale scorurilor observate și ale erorilor de măsură. Coeficientul de fidelitate reflectă doar mărimea raportului acestor variante ($\sigma^2(E_{m^*})/ \sigma^2(X_{m^*})$). Din această cauză, Standardele APA (1985) recomandă ca în manualul testului să figureze, alături de coeficientul de fidelitate al testului m, eroarea standard de măsură, $\sigma(E_{m^*})$

Coeficienții de fidelitate se vor utiliza la compararea testelor. Eroarea standard de măsură va servi la interpretarea scorurilor testului, mai precis, la determinarea intervalelor de încredere pentru scorurile reale.

Raționamentul care stă la baza procedurii de estimare a intervalului de încredere pentru scorul real al unui subiect este identic cu cel utilizat în statistică pentru estimarea intervalului de încredere al mediei populației (Traub, 1994). Se presupune că valorile variabilei X_{mp} pentru persoana $p \in \mathbf{P}$, la testul m sunt repartizate normal, cu media t_{mp} și cu abaterea standard $\sigma(E_{mp})$. Cunoscând valoarea unui scor observat pentru persoana p, x_{mp} și considerând valoarea $\sigma(E_{mp})$

egală cu eroarea standard de măsură în populația P, $\sigma(E_{m^*})$, intervalul de încredere al scorului real, corespunzător unei probabilități de 68% este

$$(x_{mp}-\sigma(E_{m^*}),\ x_{mp}+\sigma(E_{m^*}))$$

Surse de eroare în testarea psihologică

Înainte de a calcula coeficientul de fidelitate al unui test este absolut necesar să se precizeze ce reprezintă scorurile reale ale testului și care sunt sursele posibile ale erorilor de măsură. Aceasta, deoarece o anumită caracteristică a unei persoane poate fi măsurată prin scorurile reale ale unui test sau poate influența scorurile observate, ca eroare de măsură, la un alt test. Condițiile de administrare a testului (lumina, căldura, zgomotul etc.), când diferă de cele preconizate, pot modifica destul de mult scorurile observate. Din aceste motive, nu este suficient să se indice o valoare a coeficientului de fidelitate pentru a arăta cât de bine sunt reflectate scorurile reale ale testului de către scorurile observate, ci trebuie să se precizeze:

- Caracteristicile persoanelor cărora le este destinat testul (nivel de stu-dii, vârstă etc.);
- Condițiile în care se administrează testul.

Deci nu se vorbește despre fidelitatea unui instrument de măsură, ci despre fidelitatea acelui test aplicat la *o anumită populație* și în *anumite condiții*.

Au fost detectate mai multe surse ale variabilității scorurilor, deci ale consistenței sau inconsistenței performanțelor la teste. Ele au fost sintetizate în șase categorii (Tabelul 3-1).

Tabelul 3-1
Surse posibile de variabilitate care pot afecta scorurile unui anumit test

I. Particularități durabile și generale (cu influență asupra rezolvării mai multor teste) ale persoanei

A. Nivelul unor deprinderi care pot influența modul de a opera cu o serie de teste (ex. un rezolvitor de cuvinte încrucișate va fi avantajat la rezolvarea unor teste)

B. Deprinderile și tehnicile generale solicitate în testare (ex. "test-de excepție" - pentru profesioniști sau "test-pentru necunoscători")

C. Abilitatea generală de a înțelege instrucțiunile de lucru cu testele (ex. persoanele cu deficiențe intelectuale vor înțelege mai greu instrucțiunile de operare cu testul, la fel cele neobișnuite cu examinări prin teste)

II. Particularități durabile dar specifice (cu influență numai asupra acestui test) ale persoanei

A. Specifice față de testul luat ca întreg

1. Nivelul individual de dezvoltare al abilității măsurate de acest test și care nu influențează rezolvarea altor teste (ex. o persoană dominantă va obține scoruri mari la scala de *Dominanță* din *Inventarul Psihologic California*, dar nu și la scala de *Responsabilitate*)
2. Cunoștințe și deprinderi specifice formei particulare a testului (un programator va fi mai performant pe un test care solicită rezolvarea de scheme logice lacunare; un ceasornicar va avea performanțe mai bune la un test de dexteritate digitală care cere operarea cu obiecte mici)
3. *Patterne* de răspuns stabile (ex. la un test cu răspunsuri la alegere, marcarea mai frecventă a răspunsului A; la un test cu răspunsuri "Adevărat/Fals", tendința de a

alege opțiunea Adevărat în toate situațiile de indecizie etc.)

B. Specifice anumitor itemi ai testului

1. "Șansa" unei persoane să cunoască răspunsul la un anumit fapt particular cerut de un item (ex. la un test de cunoștințe tehnice, un item poate solicita o informație pe care o cunosc doar subiecții care au citit o anumită carte)
2. Tipuri de itemi cu care diferiți subiecți sunt familiarizați inegal (ex. existența în test a câtorva itemi cu răspunsuri la alegere îi poate încurca pe cei nefamiliarizați cu stilul respectiv de lucru)

III. Particularități temporare, dar generale ale persoanei (factori incidentali care afectează performanțele nu numai la acest test, ci la mai multe teste, sau chiar la toate testele, în momentul în care este administrat testul în cauză)

A. Sănătatea
B. Oboseala
C. Motivația
D. Tensiunea emoțională
E. Nivelul prea elevat al testului
F. Înțelegerea mecanismelor testării în cauză
G. Condițiile de administrare a testului, ca iluminat, căldură, umiditate, ventilație etc.

IV. Particularități temporare și specifice (cu influență numai asupra acestui test) ale persoanei

A. Specifice față de testul luat ca întreg

1. Înțelegerea sarcinii specifice cerute de testul în cauză (ex. la un test administrat de calculator, un subiect care nu a reținut modul de utilizare a tastaturii pentru formularea răspunsului poate obține un scor care nu îl caracterizează)

2. Mici trucuri sau tehnici neortodoxe de rezolvare a testului (ex. la un test de cunoștințe alcătuit din itemi cu răspunsuri la alegere, care nu penalizează răspunsurile greșite, marcarea tuturor răspunsurilor propuse la fiecare întrebare conduce la obținerea scorului maxim)
3. Nivelul de antrenare în rezolvarea unor sarcini specifice cerute de testul în cauză (intervine mai ales în operarea cu teste psihomotoare) (ex. la un test de viteză de reacție, administrat de calculator, o persoană familiarizată cu tastatura calculatorului respectiv este avantajată)
4. "Setul" sau dispoziția momentană pentru rezolvarea testului (ex. unei persoane foarte ocupate în ziua testării i se va părea prea lung Inventarul Psihologic California și s-ar putea să răspundă la întâmplare)

B. Specifice față de unii itemi ai testului

1. Fluctuații și ideosincrazii ale memoriei umane (ex. la un test de cunoștințe de geografie, un elev căruia nu îi place să opereze cu numere are șanse să nu își amintească înălțimile unor munți)
2. Fluctuații neprevăzute ale atenției sau percepției, suprapuse peste nivelul general de performanță caracteristic unei persoane (ex. la un test administrat de calculator, subiectul poate, din neatenție, să apese pe o tastă greșită și să nu observe acest lucru)

V. Factori sistematici sau de noroc care afectează administrarea testului sau evaluarea performanțelor la test

A. Condițiile de testare: utilizarea unui interval de timp fixat pentru testare, prezența unor factori de distragere a atenției, claritatea instrucției etc.
B. Efectul interacțiunii personalității, sexului sau rasei examinatorului cu persoana examinată, aceasta stimulând sau inhibând performanța

C. Inconsecvența sau părtinirea în evaluarea performanțelor

VI. Variații neexplicate
A. Norocul în găsirea răspunsului corect, atunci când subiectul ghicește răspunsul
B. Distragerea momentană a atenției

Adaptare după: Thorndike (1949) și Murphy și Davidshofer (1991).

Exemple.

I. La o serie de teste de raționament administrate unui grup de tehnicieni, cei care sunt înscriși la cursurile unei facultăți de profil, deci sunt mai obișnuiți cu munca intelectuală, sunt antrenați într-o activitate de studiu, vor obține rezultate mai bune decât colegii lor neimplicați într-un efort profesional mai intens.

II. Unii tehnicieni care au lucrat în proiectare se vor descurca mai bine în descifrarea unui test în care au fost introduși itemi axați pe reprezentări spațiale și mai slab la alte tipuri de teste. Eliminarea din test a itemilor respectivi va duce la performanțe mai mici.

III. Dacă examenul psihologic se derulează într-o încăpere friguroasă, performanțele obținute la toate testele administare vor fi mult reduse.

IV. La un test de cunoștințe de istorie antică, un elev care în ziua precedentă a vizionat filmul "Dacii" are șanse mai mari decât alții să își amintească răspunsul la unii itemi.

V. Dacă pe parcursul administrării testului în vecinătatea încăperii în care are loc examinarea se produc zgomote puternice, performanțele la testul respectiv vor fi diminuate.

VI. Prezența unui fluture în încăperea în care se efectuează o testare contracronometru poate distrage atenția unor persoane, scăzându-le astfel performanța.

Înțelegerea noțiunii de fidelitate, ca și tehnicile de estimare a fidelității scorurilor testelor, depinde în ultimă instanță atât de definirea precisă a atributelor care se măsoară, cât și de sursele de inconsistență specifice măsurării calităților respective.

Metode de estimare a fidelității

În practică, coeficienții de fidelitate pot servi unuia sau ambelor obiective: (1) să estimeze precizia măsurării realizate de o procedură particulară cum ar fi aceea a unui instrument de măsură ori test psihologic, sau (2) să estimeze consistența performanței unei proceduri de măsurare. Al doilea obiectiv îl include de fapt pe primul. Logic, este posibil să avem performanțe nefidele din partea unei persoane la un test considerat ca fiind fidel, dar obținerea de performanțe fidele la un instrument nefidel este un lucru imposibil (Wesman, 1952). Obiectivele menționate pot fi ușor surprinse în diferitele metode de estimare a fidelității. Fiecare din metodele pe care le vom discuta în continuare are în vedere unele din condițiile care pot genera modificări nesistematice în scorurile testului și în consecință pot afecta eroarea de măsură a testului.

Pe scurt, obiectivul estimării fidelității este să determine cât de mult din variabilitatea scorurilor observate la test este atribuită erorilor de măsurare și cât de mult este ea atribuită variabilității scorurilor adevărate. O strategie adecvată acestui obiectiv este metoda utilizării testelor paralele. Un exemplu de teste paralele sunt *Domino 48* și *Domino 70* (Pitariu, 1975). Ambele forme se administrează simultan la același grup de subiecți. Diferențele dintre scorurile obținute la cele două teste se datorează erorilor de măsurare. Dacă coeficientul de corelație liniară între scorurile celor două teste este apropiat de unu,

se poate trage concluzia că erorile de măsurare sunt mici, deci testele au o fidelitate bună.

Metoda testelor paralele este o soluție conceptuală ideală de estimare a fidelității, nu și una practică. Motivul este simplu: proiectarea lor este dificilă. În consecință, au fost concepute o serie de alte modele de calcul al fidelității. Literatura de specialitate recomandă utilizarea următoarelor metode de calcul a fidelității, înafara metodei formelor paralele pe care am descris-o mai sus:

- metoda test-retest - utilizată în calculul stabilității scorurilor obținute în două instanțe de testare cu aceeași probă. Ne arată cât de stabile sunt scorurile în timp.
- metoda înjumătățirii (*split-half*) - produce un coeficient de echivalență, fiind o variantă a metodei testelor paralele.
- metoda analizei consistenței interne - coeficientul de fidelitate se calculează pe baza analizei consistenței interne a unui test administrat o singură dată. Din această catgorie fac parte coeficienții α al lui Cronbach, λ_3 al lui Guttman și coeficienții Kuder-Richardson (KR-20 și KR-21)

Prezentăm în continuare (după Albu & Pitariu, 1993) tipurile de coeficienți de fidelitate, metodele prin care se obțin și condițiile în care pot fi utilizați.

Metoda test-retest

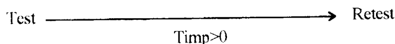
Metoda test-retest evaluează gradul în care scorurile observate la test sunt constante de la o administrare la alta.

Se procedează astfel :

- se administrează testul unui grup de persoane;
- după un interval de timp se administrează același test, acelorași indivizi;
- se calculează coeficientul de corelație liniară între scorurile observate în cele două situații.

Valoarea obținută se folosește pentru estimarea fidelității testului, considerându-se că testul este paralel cu el însuși, ceea ce înseamnă că între cele două administrări ale sale, scorurile reale ale persoanelor nu s-au schimbat și presupunând că, pentru fiecare persoană, erorile de măsură în cele două cazuri sunt variabile aleatoare independente. Coeficientul de corelație calculat prin această metodă se numește coeficient de stabilitate.

Procedura experimentală poate fi schițată astfel:



Metoda test-retest este utilă atunci când scorurile reale măsoară caracteristici *durabile generale* și *specifice* ale persoanelor.Factorii temporari (starea de oboseală sau boală a celui examinat, emoțiile, temperatura și iluminarea sălii de testare etc.) pot influența pozitiv sau negativ scorurile obținute de o persoană care participă la testare. De fapt, nu se modifică scorul real al persoanei, ci se acționează asupra erorilor de măsură.

Atât în cazul testelor de aptitudini sau personalitate, cât și în acela al testelor de cunoștințe, este posibil ca scorurile reale ale persoanelor să se modifice de la o administrare la alta. Pe de o parte, intervine uitarea, care determină o scădere a scorurilor reale. Pe de altă parte, în intervalul de timp dintre două testări, persoanele pot dobândi cunoștințe și deprinderi noi, ceea ce face să crească scorurile reale. Este motivul pentru care coeficientul de corelație liniară între scorurile observate la test și la retest nu măsoară fidelitatea testului. El arată cât de stabile în timp sunt aptitudinile, deprinderile sau cunoștințele verificate și cum se schimbă nivelul acestora la persoanele examinate.

Un coeficient de stabilitate poate avea o valoare mare atunci când :

- persoanele examinate nu s-au schimbat, în ceea ce privește particularitatea investigată, între cele două testări;
- nivelul calității investigate (aptitudine, deprindere, cunoștințe) variază, la fiecare persoană, proporțional cu valoarea detectată la prima testare;
- nivelul particularității verificate se modifică uniform (cu aceeași cantitate pentru fiecare persoană) în grupul testat.

Pentru a interpreta corect valoarea coeficientului de stabilitate, ar trebui să se calculeze, pentru fiecare persoană, diferența scorurilor observate la cele două testări, să se determine media acestor diferențe și să i se aplice un test de semnificație. Dacă se constată că media diferă semnificativ de zero, deci persoanele s-au schimbat între cele două testări, atunci determinarea curbei de regresie între scorurile la retest și cele la test ar ajuta la înțelegerea modului în care a variat nivelul particularității măsurate. De exemplu, o curbă de regresie de forma $y=x+b$ denotă o schimbare uniformă; semnul coeficientului b arată dacă este vorba despre o creștere ($b>0$) sau o descreștere ($b<0$).

Valoarea unui coeficient de stabilitate este influențată atât de lungimea intervalului de timp dintre test și retest, cât și de condițiile în care au loc cele două administrări (de exemplu, la începutul și la sfârșitul unui semestru sau la sfârșitul unui semestru și la începutul semestrului următor). Din această cauză, pentru un test nu se poate vorbi despre un singur coeficient de stabilitate, ci de mai mulți.

Dezavantajul metodei test-retest constă în faptul că pretinde două administrări ale testului, ceea ce necesită timp și cheltuieli materiale.

Metoda înjumătățirii (split-half)

Un caz particular al metodei formelor paralele și care elimină cele două deficiențe majore ale acestora (proiectarea unei forme paralele a testului și administrarea separată a celor două forme) o constituie metoda înjumătățirii sau *split-half*.

Metoda *split-half* constă în următoarele:

- se administrează testul unui grup de persoane;
- se împarte testul în jumătate;
- se calculează coeficientul de corelație între scorurile observate la cele două jumătăți.

Dacă jumătățile testului sunt măsurători paralele, atunci valoarea calculată reprezintă coeficientul de fidelitate al fiecăreia dintre părți. Cu ajutorul lui, aplicând formula Spearman-Brown, se poate determina coeficientul de fidelitate al întregului test:

$$r_{11} = \frac{2r}{1+r}$$

unde r_{11} este fidelitatea testului și r este corelația calculată între cele două jumătăți ale testului.

Exemplu. Dacă corelația între scorurile la itemii pari și scorurile la itemii impari este .80, atunci fidelitatea estimată a testului este:

$$r_{11} = \frac{2(.80)}{1+.80} = .89$$

Procedura experimentală poate fi schițată astfel:



Metoda *split-half* rezolvă multe din problemele teoretice și practice inerente metodei test-retest și metodei formelor paralele. Mai întâi, nu este necesar să se administreze două teste diferite și nici să se administreze de două ori același test. În felul acesta nu există pericolul modificării scorurilor reale între administrări și nici cel al memorării și repetării răspunsurilor date la primul test.

Dificultatea operării cu metoda *split-half* constă în construirea jumătăților care să fie măsurători paralele. Se recomandă (Cronbach, 1943; Pitariu, 1978) să se elaboreze perechi de itemi care să fie similari ca :

- formă (mod de prezentare a textului, a ilustrațiilor);
- conținut;
- dificultate;
- medie și abatere standard a scorurilor;
- tip de întrebare (cu răspuns liber, cu răspuns la alegere etc.).

De asemenea, itemii trebuie să fie independenți, adică unul să nu sugereze răspunsul la celălalt.

Din fiecare pereche, un item va intra într-o jumătate de test, iar al doilea, în cealaltă jumătate.
Există multe posibilități de ordonare a itemilor în cadrul testului. În Tabelul 3-2 sunt ilustrate 10 variante de înjumătățire a unui test compus din șase itemi (itemii sunt considerați paraleli între ei).

Tabelul 3-2
Zece modalități posibile de înjumătățire a unui test format din șase itemi (după Murphy & Davidshofer, 1991)

Itemi de test jumătatea A	Itemi de test jumătatea B	Fidelitatea estimată
1, 2, 3	4, 5, 6	.64
1, 2, 4	3, 5, 6	.68
1, 2, 5	3, 4, 6	.82
1, 2, 6	3, 4, 5	.79
1, 3, 4	2, 5, 6	.88
1, 4, 5	2, 3, 6	.81
1, 5, 6	2, 3, 4	.82
2, 3, 5	1, 4, 6	.72
2, 4, 5	1, 3, 6	.71
2, 4, 6	1, 3, 5	.74

Ceea ce se poate observa în Tabelul 3-2 este că fiecare din cele zece posibilități de grupare produce un alt coeficient de fidelitate. Cu alte cuvinte, corelația dintre A și B variază după cum este testul înjumătățit. Problema este, care din coeficienții de fidelitate obținuți este cel real? Singurul răspuns posibil este că metoda split-half nu este una perfectă și că nu găsim nici un argument în favoarea unei metode de înjumătățire față de alta.

Alte precizări. Nu se recomandă plasarea celor două părți una după cealaltă, întrucât intervine oboseala persoanelor examinate, care va influența scorurile la cea de a doua jumătate, iar dacă testul este administrat cu limită de timp, este foarte probabil ca mulți itemi din cea de a doua parte să nu fie rezolvați datorită timpului de lucru insuficient și nu din cauza absenței aptitudinilor, deprinderilor sau cunoștințelor cerute de întrebări. Pentru a evita asemenea situații, se

poate adopta metoda par-impar : itemii unei jumătăți vor ocupa locurile cu număr de ordine par, iar ceilalți, locurile cu număr de ordine impar. Se obișnuiește să se plaseze itemii în ordinea crescătoare a dificultății lor (Lord & Novick, 1968).

Apare însă o problemă în cazul când testul se administrează cu limită de timp. Fiecare persoană care nu reușește să răspundă la ultimii 2r itemi va avea scorul zero la aceștia și, deci, scorurile la itemii pari și la cei impari vor apărea mai puternic corelate decât sunt în realitate. Atunci, coeficientul de corelație al scorurilor părților ar putea să difere mult de coeficientul de fidelitate al unei jumătăți a testului.

În concluzie, metoda *split-half* nu se va aplica atunci când administrarea testului se face cu limită de timp.

Dacă testul este alcătuit din două părți paralele, coeficientul de fidelitate al testului poate fi calculat prin oricare din formulele Spearman-Brown, Guttman, Flanagan sau Rulon (Albu & Pitariu, 1993; Pitariu, 1976).

Metoda analizei consistenței interne

Această metodă utilizează pentru estimarea fidelității unui test variantele și covariantele scorurilor observate ale itemilor. Este avantajoasă, necesitând numai o singură administrare a testului. Pe baza scorurilor observate ale itemilor se calculează de obicei unul dintre următorii coeficienți :

- coeficientul α al lui Cronbach;
- coeficientul r_{20} (KR-20) sau r_{21} (KR-21) al lui Kuder-Richardson (dacă itemii testului sunt binari);
- coeficientul λ_3 al lui Guttman.

Acești coeficienți sunt denumiți *coeficienți de consistență internă;* uneori îi mai găsim și sub denumirea de *coeficienți de omogenitate.* Ei indică măsura în care itemii testului sunt intercorelați.

În general, pentru a calcula coeficienții de consistență internă se procedează astfel:

- se administrează testul unui grup de indivizi;
- se utilizează formula aleasă pentru a calcula coeficientul de fidelitate.

Menționăm că, în general, metodele de estimare a consistenței interne se aseamănă într-un anumit fel cu metoda *split-half.* Dacă metoda *split-half* compară o jumătate a testului cu cealaltă jumătate, în estimarea consistenței interne se compară fiecare item cu fiecare alt item.

Formule pentru calcularea coeficientului de consistență internă:
Procedura cea mai populară pentru calculul consistenței interne a fost elaborată de Kuder și Richardson (1937). Ea este cunoscută sub denumirea de formula 20 a lui "Kuder-Richardson" sau, simplu, KR-20. Această formulă are următoarea structură:

$$r_{20} = \left(\frac{n}{n-1}\right)\frac{\sigma^2 - \sum_{i=1}^{n} p_i q_i}{\sigma^2}$$

unde r_{20} este coeficientul de fidelitate al testului, n reprezintă numărul itemilor din test și σ abaterea standard a scorului total al testului. p_i este frecvența relativă a celor care au răspuns corect la itemul i și q_i este frecvența relativă a celor care au răspuns necorespunzător la itemul $i (q_i = 1-p_i)$.

KR-21 este un caz particular al formulei KR-20 în care se presupune că toți itemii testului au același grad de dificultate. În acest caz formula este:

$$r_{21} = \left(\frac{n}{n-1}\right)\frac{\sigma^2 - n\overline{pq}}{\sigma^2}$$

unde $\overline{p}$ și $\overline{q}$ sunt mediile frecvențelor relative ale răspunsurilor corecte, respectiv greșite, pentru cei n itemi.

Coeficientul α al lui Cronbach (Cronbach, 1943) este o generalizare a formulei KR-20, fiind aplicabil și atunci când testul conține itemi care prezintă o varietate mai mare de răspunsuri. Un chestionar de atitudini poate, de pildă, să pretindă cotarea unui item pe o scală de tipul: "obișnuit=4", "uneori=3", "rar=2", "niciodată=1". Pentru astfel de situații, a fost elaborată o formulă , cunoscută sub denumirea de coeficientul alpha (Cronbach, 1951; Novick & Lewis, 1967). În această formulă, valoarea Σpq din KR-20 este înlocuită cu , suma variantelor scorurilor itemilor. Formula coeficientului α este următoarea:

$$\alpha = \left(\frac{n}{n-1}\right)\frac{\sigma^2 - \sum_{i=1}^{n}\sigma_i^2}{\sigma^2}$$

Coeficientul λ_3 al lui Guttman, aduce o notă de precizie suplimentară prin faptul că el este cel mai apropiat de valoarea reală a coeficientului de fidelitate al testului și poate lua valori pozitive atunci când ceilalți coeficienți de consistență internă sunt negativi și, astfel, sunt lipsiți de utilitate.

Formula de calcul este următoarea:

$$\lambda_3 = 1 - \frac{\sum_{i=1}^{n}\sigma_i^2}{\sigma_x^2} + \frac{1}{\sigma_x^2}\sqrt{\frac{n}{n-1}\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n}\sigma^2(Y_i, Y_j)}$$

σ_x^2 este varianța scorurilor observate ale testului, iar σ_i^2 reprezintă dispersia scorurilor observate la itemul i, Y_i, $i=1,2,...n$. $\sigma(Y_i, Y_j)$ este covarianța scorurilor observate la itemii i și j, $i,j=1,2,...,n$ și $i\neq j$.

Interpretarea coeficienților de fidelitate

Dintr-o mulțime de teste având toate celelalte caracteristici egale, este de preferat testul care are cea mai mare fidelitate. Dar, pe de o parte, testul cel mai fidel poate fi și cel mai scump sau mai dificil de administrat. Pe de altă parte, caracteristicile testelor sunt rareori egale.

Este imposibil să se specifice un nivel minim care să fie impus coeficienților de fidelitate ai tuturor testelor. Murphy și Davidshofer (1987) formulează următoarele recomandări:

- se vor cere niveluri înalte de fidelitate testelor folosite pentru a se lua decizii importante asupra persoanelor și celor care împart indivizii în mai multe categorii pe baza unor diferențe relativ mici între ei;
- se vor accepta niveluri scăzute de fidelitate atunci când testele se utilizează pentru a lua decizii preliminare, nu finale, și în cazul testelor care împart persoanele într-un număr mic de categorii, pe baza unor diferențe individuale mari.

T.L. Kelley (Guilford, 1965) sugerează că, în general, pentru a discrimina bine între indivizi, un test trebuie să aibă un coeficient de fidelitate de cel puțin .94. Unii au admis valoarea minimă de .90.

Există, însă, mai multe teste foarte utile cu coeficienți de fidelitate mai mici de 0.80. În practică ne așteptăm la coeficienți de fidelitate cuprinși între .70 și .98 (Guilford, 1965).

VALIDITATEA

Câteva noțiuni utile

Problemele de psihologie a personalului sunt complexe, noțiunile utilizate le întâlnim în lucrări cu caracter general de psihologia muncii și organizațională, de psihodiagnostic ori de management al resurselor umane. Pentru ca cele ce urmează să fie mai ușor înțelese, dorim să oferim câteva explicații recapitulative, pentru câteva noțiuni pe care le utilizăm în construcția și analiza testelor psihologice.

Constructul psihologic. Un construct este, în general, o idee "construită" de experți dintr-un domeniu particular al științei, pentru a rezuma un grup de fenomene sau de obiecte. După Murphy și Davidshofer (1991), orice construct are două proprietăți:

- este o abstractizare a unor regularități din natură;
- nu este observabil direct, dar poate fi conectat cu entități sau evenimente concrete, observabile.

Un construct psihologic, arată J.F. Binning și G.V. Barett (Silva, 1993), este "o etichetă pentru o grupare de comportamente care co-variază". Pentru Standardele APA din 1985, constructul este " o noțiune teoretică derivată din cercetare și din alte experiențe, care a fost construită pentru a explica patternuri de comportament observabile".

Faptul că trăsăturile psihice sunt privite uneori ca "patternuri de comportamente" (care se manifestă aproape la fel în situații diferite) (Endler & Magnusson, 1976) poate genera confuzii între constructele psihologice și trăsăturile psihice. Deosebirea dintre ele a fost

explicată de J. Loevinger astfel: "Trăsăturile există în indivizi. Constructele există în mințile și în revistele psihologilor. Este adevărat că psihologii nu cunosc niciodată trăsăturile direct, ci numai prin intermediul constructelor lor, dar datele ce trebuiesc judecate sunt manifestări ale trăsăturilor și nu manifestări ale constructelor" (Silva, 1993).

Constructele create pentru interpretarea rezultatelor testelor psihologice sunt identificate, de obicei, cu câte o clasă de răspunsuri (L. J. Cronbach, după Silva, 1993).

Fiecare construct psihologic este fundamentat pe o teorie, care permite descrierea și predicția comportamentelor legate de el, în situații specifice.

Pentru elaborarea unui test care să îl măsoare, constructului i se asociază o variabilă cantitativă despre care se presupune că se află în relație crescătoare cu scorurile testului. Se formulează apoi o serie de afirmații referitoare la modul în care se comportă persoanele care au valori mari ale acestei variabile, comparativ cu cele care au valori scăzute. De exemplu, vrând să dezvolte un test de evaluare a motivației indivizilor pentru propria realizare, McClelland și colaboratorii săi au formulat următoarele descrieri de comportamente (Thorndike & Hagen, 1961):

- Persoanele care au o motivație puternică pentru realizare au rezultate școlare bune.
- Motivația pentru realizare este mai mare dacă subiecților li se administrează în prealabil teste despre care li se spune că măsoară inteligența.
- Cei care au o motivație puternică pentru realizare completează mai mulți itemi într-un test de viteză a cărui funcție le-a fost explicată.
- Motivația pentru realizare este mai mare la copiii a căror familie le-a încurajat de timpuriu independența.

Fiecare dintre aceste descrieri este bazată pe un fel de "teorie a comportamentului uman". De exemplu, realizarea şcolară este văzută ca o combinaţie între aptitudinea şcolară şi efort. Dintre persoanele cu acelaşi nivel aptitudinal, cele care au o motivaţie mai mare pentru realizare depun un efort mai mare şi obţin rezultate şcolare mai bune (Thorndike & Haagen, 1961).

Comportamentele prin care autorul testului consideră că se manifestă constructul pot fi influenţate de diverse variabile, precum vârsta, sexul, educaţia primită etc. Pentru unele dintre acestea legătura cu constructul poate corespunde unei situaţii reale, observată în populaţia căreia i se adresează testul şi acceptată de teoria care stă la baza constructului. Pentru altele, însă, ea denotă că aceste comportamente nu sunt caracteristice constructului. De asemenea, comportamentele prin care este descris constructul măsurat de test se pot asocia cu comportamente prin care se manifestă alte constructe psihologice, evaluate prin teste existente.

O descriere amănunţită a constructului trebuie să conţină şi relaţiile acestuia cu cât mai multe variabile (fiziologice, psihologice etc.). De exemplu, dacă se urmăreşte realizarea unui test pentru evaluarea "inteligenţei", se pot avea în vedere următoarele relaţii ale constructului "inteligenţă" cu alte variabile (Gronlund, 1967):

- Scorurile testului cresc cu vârsta (se presupune că inteligenţa creşte cu vârsta aproximativ până la 16 ani).
- Scorurile testului prezic succesul şcolar.
- Scorurile testului concordă cu aprecierile făcute de profesori despre inteligenţa elevilor.
- Scorurile testului corelează pozitiv cu scorurile unor teste aşa-numite "de inteligenţă".
- Scorurile testului discriminează între un grup de copii handicapaţi mintal şi unul format din copii capabili de performanţe superioare.

- Scorurile testului sunt influenţate doar puţin de procesul instrucţional urmat de subiect.

Fiecare dintre afirmaţiile formulate în legătură cu constructul pe care testul trebuie să îl măsoare urmează să fie verificată experimental. Dacă unele dintre ele nu se confirmă, trebuie revăzută teoria care stă la baza constructului, pentru a sesiza eventualele omisiuni sau neconcordanţe.

Domeniul de conţinut măsurat de un test. Prin "domeniu de conţinut al unui test" se înţelege "mulţimea tuturor comportamentelor care pot fi utilizate pentru a măsura atributul specific sau caracteristica la care se referă testul" (Murphy & Davidshofer, 1991).

Definirea domeniului de conţinut este asemănătoare celei de definire a unui construct. În ambele cazuri, autorul testului, pe baza ideii pe care şi-a format-o despre ceea ce vrea să măsoare, reţine acele comportamente care presupune că ar fi manifestări ale constructului, respectiv expresii ale performanţei sau ale cunoştinţelor subiecţilor.

Deosebirea dintre un construct şi un domeniu de conţinut constă numai în gradul de abstractizare implicat şi adesea este dificil de făcut distincţie între ele (Guion, 1986).

Descrierea domeniului de conţinut trebuie în primul rând să precizeze limitele acestuia. Apoi, pentru fiecare componentă se apreciază importanţa. În cazul testelor destinate să măsoare rezultatele unui program de instruire (cunoştinţe sau deprinderi), descrierea domeniului de conţinut a testului se bazează pe analiza conţinutului şi a obiectivelor instrucţionale ale cursului urmat de persoanele cărora li se adresează testul. Se încearcă să se includă în domeniu aspectele pe care mai mulţi experţi în domeniu le consideră importante.

Exemplu.

Un test de cunoștințe de istorie, administrat elevilor din clasa a VIII-a la sfârșitul celui de-al doilea trimestru, poate avea ca domeniu de conținut "cunoașterea istoriei României, conform materiei predate în cadrul lecțiilor".

Descrierea domeniului de conținut cuprinde:

a. limitele domeniului: se au în vedere cunoștințele cuprinse în lecțiile 11-18 din manualul de clasa a VIII-a;

b. componentele domeniului:

- probleme: sociale, politice, culturale;
- zone: Moldova,Țara Românească, Transilvania;
- timp: secolul al XVI-lea și al XVII-lea.

c. importanța relativă a componentelor domeniului: este indicată în Tabelul 3-3. Procentele se referă la timpul alocat fiecărei probleme în cadrul lecțiilor predate la clasă.

Tabelul 3-3
Importanța componentelor domeniului de conținut pentru un test de cunoștințe de istorie

Zona	*Probleme*		
	sociale	politice	culturale
Moldova	10%	10%	10%
Țara Românească	8%	12%	15%
Transilvania	13%	17%	5%

Există, însă, multe domenii de conținut care nu pot fi descrise detaliat după modelul prezentat. De exemplu, domeniul de conținut al unui test de aptitudini cognitive poate fi format din itemi extrași din teste cunoscute, care se consideră că evaluează aptitudinea respectivă. Într-o asemenea situație este dificil de precizat care sunt limitele domeniului de conținut.

Criteriul. În teoria testelor psihologice, prin criteriu se înțelege "o variabilă pe care încercăm să o prevedem în general cu ajutorul testelor" (Grand dictionnaire de la psychologie, 1994) sau "o măsură acceptată a comportamentului evaluat de test" (Corsini, 1994). În aprecierea personalului, criteriul este o "măsură a nivelului de performanță exprimată în termeni cantitativi, bazată pe o descriere completă a muncii prestate" (Corsini, 1994).

Se recurge la utilizarea criteriilor atunci când se dorește să se analizeze funcția unui test, de exemplu pentru a verifica dacă acesta măsoară același lucru ca și un alt test, cunoscut, sau dacă permite să se prevadă succesul într-o anumită activitate.

Pentru a face posibilă aprecierea corectă a calităților testului cercetat, criteriul trebuie să îndeplinească o serie de condiții. Amintim câteva dintre acestea (adaptare după Corsini, 1994):

- Să fie exprimat printr-o variabilă care ia valori dintr-o mulțime total ordonată. De obicei se preferă criteriile cu valori cantitative.
- Să fie într-o legătură strânsă cu variabila la care se referă testul: să măsoare corect variabila respectivă (să fie el însuși un test valid și fidel) sau să evalueze corect performanțele profesionale, luând în considerare diversele dimensiuni ale acesteia.
- Să reflecte comportamentul sau performanța persoanelor când acestea nu se află sub acțiunea unor factori perturbatori (de exemplu, să nu aprecieze că o persoană are o productivitate scăzută în muncă dacă în perioada respectivă mașina pe care aceasta lucra a fost mult timp stricată).
- Să fie exprimat în aceleași unități sau în unități comparabile pentru toate persoanele (de exemplu, în cote standardizate obținute prin raportare la etaloane diferite în funcție de sex).
- Atunci când criteriul este multidimensional -ceea ce se întâmplă adesea în aprecierea personalului-, valoarea criteriului pentru fiecare persoană să reflecte atât importanta diverselor aspecte, cât și nivelul individului la fiecare dintre ele.

Anastasi (1954) enumeră câteva criterii utilizate în practica psihologică pentru analizarea testelor care măsoară constructe.

Acestea sunt:

1. Vârsta: se folosește drept criteriu pentru testele de inteligență, întrucât se consideră că scorurile acestora trebuie să crească pe parcursul copilăriei, până la maturitate. Nu este un criteriu bun pentru testele care măsoară funcțiuni care nu se schimbă cu trecerea anilor, cum sunt variabilele de personalitate.

2. Cunoștințele școlare, exprimate prin note școlare, rezultate la teste de cunoștințe, evaluări făcute de profesori asupra "inteligenței" elevilor: sunt adesea criterii pentru testele care măsoară aptitudinea școlară.

3. Performanțele la diverse programe de instruire specială, cum sunt cursurile de dactilografie, programare, muzică etc., furnizează criterii pentru teste de aptitudini speciale.

4. Performanța profesională servește uneori drept criteriu pentru teste de personalitate. Performanța în muncă este un criteriu multidimensional. În compoziția sa intră aspecte ce țin de (Gilmer, 1971; Pitariu, 1994):

- cantitatea realizată în unitatea de timp (piese produse, cuvinte dac-tilografiate, itemi codificați etc.);
- calitatea muncii (număr de piese rebutate, număr de erori de codificare etc.);
- timpul de muncă neutilizat (absențe, întârzieri, pauze neautorizate etc.);
- stabilitatea în muncă (vechimea în muncă, schimbări de profesie, transferări, demisii etc.);
- costul necesar pentru instruire (timpul de instruire pentru realizarea unei producții standard, costul materialului stricat în timpul instruirii, viteza de avansare etc.);
- cunoștințe profesionale;

- satisfacția în muncă (număr de revendicări, de sugestii, examinări psihiatrice etc.);
- evaluări făcute de superiori sau de colegi.

5. Evaluările (făcute de profesori, instructori, superiori) sunt folosite mai mult pentru a obține informații despre alte criterii, cum ar fi cunoștințele școlare, performanța la diverse cursuri, reușita profesională. Au rol de criteriu pentru unele teste care măsoară trăsături psihice, precum anxietatea, onestitatea, originalitatea etc.

6. Rezultatele altor teste se utilizează ca și criteriu atunci când se urmărește să se construiască un test care să măsoare același lucru ca și criteriul dar să fie mai ușor de administrat sau/și de cotat ori să fie mai ieftin. De exemplu, scala *Stanford-Binet* a servit drept criteriu pentru multe teste de inteligență.

7. Grupele contrastante reprezintă de obicei un criteriu compozit, care cuprinde aspecte diverse din viața zilnică. De exemplu, pentru un test de inteligență grupele contrastante pot fi constituite dintr-o clasă de copii proveniți de la o școală de debili mintali și o clasă de elevi de aceeași vârstă dintr-o școală obișnuită. Criteriul pe baza căruia s-au format grupele este alcătuit din ansamblul de factori care au determinat cuprinderea copiilor în școala ajutătoare. Grupele contrastante se folosesc frecvent ca și criteriu pentru testele de personalitate. De exemplu, pentru evaluarea unor trăsături legate de activitatea socială se poate forma o grupă de studenți cu multiple preocupări extrașcolare și una cu studenți fără asemenea preocupări.

Adesea, pentru un același test se găsesc mai multe subcriterii, fiecare referindu-se la un alt aspect. Așa se întâmplă, de pildă, când testul trebuie să prevadă reușita profesională.

În asemenea cazuri există două posibilități de a acționa pentru a le lua în considerare pe toate:

1. Se combină subcriteriile într-un criteriu compus sau global; pe baza scorurilor subcriteriilor se calculează pentru fiecare persoană o cotă unică.
2. Se reunesc subcriteriile într-un criteriu multiplu: fiecare subcriteriu are o valoare particulară și este utilizat separat în aprecierea testului.

Prima variantă de lucru este mai convenabilă în privința prelucrării și a interpretării datelor, întrucât fiecare persoană are un singur scor, care permite ierarhizarea candidaților. Pentru aplicarea sa, este necesar, însă, ca scorurile subcriteriilor să fie exprimate în aceleași unități de măsură: în scoruri standardizate, în ranguri, în unități de timp sau în costuri.

De exemplu, pentru munca de operator la calculator se pot fixa două variabile criteriu:

- numărul de formulare de date introduse într-o oră;
- numărul de erori comise la introducerea datelor din aceste formulare.

J. Otis (Brogden & Taylor, 1949) a propus să se măsoare cele două subcriterii în unități de timp, respectiv prin:

- timpul necesar pentru introducerea unui număr fixat, N, de formulare de date;
- timpul necesar pentru depistarea și corectarea erorilor produse la introducerea datelor din cele N formulare.

Presupunerea implicită a acestui procedeu de lucru este că numărul de ore/om necesare pentru a realiza o unitate de produs este o măsură a contribuției individului la realizarea eficienței unității în care muncește.

O altă posibilitate ar fi să se măsoare toate subcriteriile prin costuri (de exemplu, costul producției realizate, costul corectării erorilor, costul supravegherii lucrătorului etc.) (Brogden & Taylor, 1949).

Pornind de la scorurile subcriteriilor se calculează, pentru fiecare persoană, o cotă, pe baza căreia se realizează un clasament al subiecților. Cei cu scorurile cele mai mari vor constitui "grupa celor cu succes profesional", iar cei cu scorurile cele mai mici, "grupa celor cu eșec profesional".

Din punct de vedere matematic se rezolvă aici o problemă de agregare a mai multor clasamente: ierarhiile de pornire (corespunzătoare scorurilor subcriteriilor) se înlocuiesc cu un singur clasament. În literatura de specialitate sunt descrise mai multe procedee de agregare a clasamentelor. Unele calculează cota unui subiect prin însumarea rangurilor ocupate de acesta în ierarhiile de pornire (Kendall, 1948), alte metode -de exemplu, metoda Borda (Păun, 1987)- acordă punctaje rangurilor și însumează punctele cuvenite subiectului în fiecare ierarhie.

În practica selecției profesionale se obișnuiește să se combine liniar, ponderat sau neponderat, scorurile subcriteriilor pentru a obține cota globală. În privința modului de stabilire a ponderilor, părerile sunt diverse:

- după H. Toops (Brogden & Taylor, 1949) ponderile trebuie să reflecte importanța fiecărui subcriteriu, apreciată de specialiști;
- M.W. Richardson (Brogden & Taylor, 1949) indică un algoritm de determinare a ponderilor astfel încât fiecare subcriteriu să aibă aceeași contribuție la covarianța scorului compus;
- M.L. Blum și J.C. Naylor (1968) propun ca ponderile să fie pro-porționale cu fidelitatea subcriteriilor.

Procedeele de agregare a clasamentelor care însumează scorurile ponderate ale subcriteriilor prezintă un neajuns atunci când ierarhiile corespunzătoare subcriteriilor nu concordă între ele. Astfel, este posibil ca pe primele locuri ale ierarhiei finale să ajungă persoane care au obținut scoruri modeste la mai multe subcriterii, dacă ele s-au aflat pe locuri fruntașe în clasamentele corespunzătoare subcriteriilor cărora li se atribuie ponderi mari.

Acest inconvenient poate avea urmări grave dacă subcriteriile nu au un caracter compensatoriu. Atunci când existența unor cunoștințe sau deprinderi este absolut necesară desfășurării cu succes a activității pentru care se face selecția și nu poate fi compensată prin alte calități ale persoanei, de asemenea necesare postului, aceste cunoștințe sau deprinderi trebuie să constituie un subcriteriu care nu va putea fi, sub nici o formă, inclus într-un criteriu compus (Landy, 1985). În asemenea situații trebuie să se utilizeze un criteriu multiplu.

Definiția validității

Când se operează cu un test psihologic sau de cunoștințe ori cu un chestionar de personalitate, trebuie să se determine două lucruri esențiale: dacă testul în cauză măsoară ceea ce pretinde că face și dacă testul poate fi utilizat în luarea unor decizii corecte. De pildă dacă o companie este interesată în deschiderea unei sucursale într-un oraș și dorește să-și selecționeze o echipă managerială, ea poate include, pe lângă alte exigențe legate de procedura de angajare, și un examen psihologic. Firește există multe strategii de selecție a managerilor, dar beneficiarii trebuie să se intereseze în primul rând dacă acestea într-adevăr măsoară ceea ce ei doresc și cât de corecte sunt deciziile bazate pe ele. Aceste două aspecte importante pe care trebuie să le conțină un test pentru a deveni operațional sunt integrate în conceptul de validitate.

Standardele APA (1954) propun patru categorii esențiale de definire a validității denumite cele "patru fețe ale validității" și care în prezent au devenit patru strategii ale validării inferențelor făcute pe baza scorurilor de test (APA Standards, 1985):

1. validitatea de conținut;
2. validitatea conceptuală;
3. validitatea predictivă;
4. validitatea concurentă.

Validitatea unui test, arată Silva (1993), se referă la cât de potrivite sunt interpretările descriptive, explicative sau predictive care se dau scorurilor sale (Silva, 1993). Într-o accepțiune actuală, conceptului de validitate i s-a dat o semnificație nouă, prin aceasta înțelegându-se "măsura în care o anumită inferență făcută plecând de la test are înțeles și este potrivită" (APA Standards, 1985).

Validarea unui test este procesul prin care se investighează gradul de validitate a interpretării propuse de acesta (APA Standards, 1985). Iată deci că validarea nu se mai referă atât la instrumentul în sine, ci la interpretarea datelor furnizate de el. Dar, pentru o interpretare corectă a scorurilor unui test este necesar să fie îndeplinite două condiții (Albu, 1995):

- să se știe exact ce reprezintă scorurile testului;
- măsurările făcute de test să fie corecte.

Aceasta ar însemna că:

- Scorurile unui test pot fi măsurători ale unei variabile construite de psihologi (construct psihologic) sau pot fi niște valori care trebuie să se găsească într-un anumit tip de relație (de obicei, relație liniară) cu rezultatele unei alte măsurători.

- Evaluările făcute de test sunt corecte dacă ierarhia subiecților realizată pe baza scorurilor testului este foarte asemănătoare clasamentului acestora pe baza variabilei măsurate de test (ale cărei valori sunt însă necunoscute) sau pe baza rezultatelor măsurătorii cu care se compară testul.

Toate verificările amintite, care se efectuează asupra testului, sunt parte integrantă a procesului de validare. Deci, validarea studiază și corectitudinea proiectării testului. Din această perspectivă, expresia "validitatea testului" devine o abreviere convențională pentru "validitatea testului și validitatea inferențelor făcute pe baza scorurilor sale".

Standardele APA (1985) apreciază că nu se poate vorbi despre "tipuri" diferite de validitate, cele recunoscute până acum sunt doar simple etichete atribuite unor categorii de strategii de validare care nu sunt disjuncte. De exemplu, validarea relativă la criteriu poate fi efectuată atât în cazul testelor care măsoară un construct, cât și în cel al testelor utilizate în predicție (Guion, 1986). În primul caz, criteriul este un construct ales datorită legăturilor pe care se presupune că le are (sau că nu le are) cu constructul măsurat de test. În cel de-al doilea caz, criteriul este o măsură a performanței sau a comportamentului în muncă.

C. L. Lawshe propune înlocuirea denumirii de "tip de validitate" cu cea de "tip de analiză a validității". Prin "analiza validității" el înțelege o procedură, un proces sau o strategie prin care se colectează sau se generează date în vederea determinării extinderii ori a robusteței inferențelor care pot fi făcute pe baza unei mulțimi de scoruri de test (Landy, 1986).

În legătură cu validitatea testelor, se impun patru idei (Hammond, 1995):

1. Există numeroase metode pentru validarea testelor, care pot fi privite ca moduri diferite de formulare a cerințelor referitoare la teste. Corespunzător scopurilor pentru care se utilizează un test, se alege strategia de validare corespunzătoare. Ea se poate baza pe analiza de itemi, pe analiză factorializă, pe determinarea corelațiilor testului cu diverse criterii etc.

2. Validitatea unui test nu poate fi estimată printr-un singur coeficient, ci se deduce din acumularea dovezilor empirice și conceptuale. Reținem că, așa cum subliniază Standardele APA din 1974, "validitatea nu se măsoară, ci se deduce". Manualul testului trebuie să conțină coeficienții de validitate, dar în cazul unei aplicări particulare a testului nu se va lua în considerare un singur coeficient, ci prin analiza lor globală se va aprecia dacă utilizarea testului este "adecvată" sau "nesatisfăcătoare". Pentru a sublinia ideea că este greșit să se spună "testul are validitate (de un anumit tip)", Standardele APA din 1985 modifică terminologia utilizată în legătură cu validitatea. Astfel, ele înlocuiesc denumirile cunoscute ale tipurilor de validitate cu următoarele "etichete" atribuite strategiilor de validare: "dovezi relative la conținut ale validității", "dovezi relative la construct ale validității" și "dovezi relative la criteriu ale validității". Prin urmare, faptul că scorurile testului folosit ca predictor corelează liniar semnificativ, la un prag de probabilitate fixat, cu valorile unui criteriu nu reprezintă o evaluare a validității relative la criteriu și nu permite să se afirme că "testul are validitate relativă la criteriu", ci doar constituie una dintre dovezile validității relative la criteriu a testului.

3. Validarea unui test este cumulativă. Procedura de validare trebuie reluată periodic pe eșantioane diferite de subiecți. Ori de câte ori se face o modificare importantă într-un test în ceea ce privește formatul,

limbajul, conținutul sau instrucțiunile de administrare, testul trebuie revalidat.

4. Validitatea unui test depinde mai mult de felul în care este folosit testul, decât de testul în sine. Tabelul 3-4 prezintă un exemplu referitor la tipurile de validitate care trebuiesc verificate pentru un singur test, compus din exerciții de aritmetică, dacă acesta este utilizat în scopuri diferite (Anastasi, 1976).

Tabelul 3-4
Diferite moduri de verificare a validității unui test, în funcție de obiectivul utilizării acestuia (Anastasi, 1976)

În ce scop este utilizat testul	Ce întrebare își pune utilizatorul testului când vrea să interpreteze scorul unui subiect	Ce tip de validitate este verificată
Verificarea cunoștințelor de aritmetică în școala primară	Cât de mult a învățat elevul în trecut?	Validitatea de conținut
Testarea aptitudinii matematice pentru predicția performanțelor în gimnaziu	Cât de bine va învăța elevul în viitor?	Validitatea de conținut
Diagnosticarea lipsei deprinderii de a face calcule	Performanța elevului indică lipsa deprinderii de a face calcule?	Validitatea relativă la criteriu (validitatea concurentă)
Măsurarea raționamentului logic	Cum poate fi descris modul de a raționa al elevului?	Validitatea conceptuală

Este absolut necesar ca atunci când se vorbește despre validitatea unui test să se precizeze despre ce validitate este vorba (predictivă, de

conţinut etc.).Dacă în cadrul unei populaţii există subgrupuri cu caracteristici diferite, este posibil ca între acestea să apară deosebiri şi în privinţa validităţii predictive a testelor. De exemplu, un test poate fi un bun predictor al performanţei numai pentru bărbaţi, nu şi pentru femei. Sau numai pentru persoane care au studii superioare etc. O variabilă care modifică validitatea predictivă a unui test, cum sunt variabilele din exemplele anterioare, se numeşte variabilă moderatoare (Dunnette, 1966).

În continuare vom prezenta "tipurile" de validitate. Atragem însă atenţia asupra a două observaţii importante făcute de Standardele APA din 1985:

- prin "tipuri" de validitate nu se înţelege "categorii disjuncte";
- nu se poate afirma că un tip de validitate este mai potrivit decât altul pentru o anumită utilizare a testelor sau pentru o categorie specifică de inferenţe bazate pe scoruri.

Validitatea testului relativă la constructul măsurat

În psihologie sunt măsurate prin intermediul testelor psihologice atribute abstracte ca inteligenţa, motivaţia agresivitatea, depresia etc., denumite constructe. Ele nu există în sens fizic, este imposibil să oferim un metru de fericire sau să identificăm un kilogram de depresie. Asemenea variabile sunt construite pornind de la fapte observabile, printr-un şir de raţionamente bazate pe ipoteze şi deducţii. Din acest motiv, validitatea relativă la construct este adesea denumită si "validitate ipotetico-deductivă" (Bacher, 1981). Toate constructele sunt însă conectate de realitate, reprezintă aspecte sau evenimente observabile. Constructele psihologice sunt evidenţiate, direct sau indirect, de comportament sau în urma conducerii unor experimente dedicate măsurării lor.

Se impune o precizare. Deoarece termenii "construct" şi "concept" sau "noţiune" sunt folosiţi adesea ca sinonimi, "validitatea relativă la construct a testului" (construct validity) este denumită şi "validitate

conceptuală" (concept validity). Între cei doi termeni există însă o deosebire, evidenţiată de L. J. Cronbach : "constructul" este "o categorie creată intenţionat pentru a organiza experienţa în enunţuri generale care se prezintă sub formă de legi" (Silva, 1993). Un construct este, deci, o noţiune elaborată şi utilizată într-un cadru ştiinţific (teoretic, metodologic sau aplicativ).

Validarea testului relativă la constructul măsurat se ocupă de calităţile psihice care contribuie la formarea scorurilor acestuia şi urmăreşte înţelegerea dimensiunilor evaluate de test. Privind constructul ca bază interpretativă a răspunsurilor subiecţilor, acest tip de validare are şi scopul de a identifica şi analiza procesele psihice declanşate/detectate de test (Silva, 1993).

L. J. Cronbach, reluând ideea formulată de Standardele APA din 1954 - conform căreia prin validarea testului relativă la constructul măsurat se validează atât testul cât şi ipotezele aflate la baza sa - scrie: "Validitatea testului şi validitatea constructului sunt inseparabile. Când un test nou este elaborat pentru măsurarea unui construct cunoscut, riscul ca validitatea să fie nesatisfăcătoare este mai mare pentru test decât pentru construct. Dar, poate să se constate şi necesitatea revizuirii constructului. Un exemplu în acest sens îl constituie abandonarea ideii tradiţonale asupra debilităţii mintale" (Silva, 1993).

Numai înţelegând constructul măsurat de test se pot formula ipoteze asupra predicţiilor pe care le face testul şi asupra relevanţei şi a reprezentativităţii conţinutului său. Deci, validarea constructului trebuie să preceadă validarea conţinutului testului şi validarea relativă la criteriu.

Pentru a putea verifica dacă un test măsoară bine un anumit construct, este necesar să se realizeze o descriere a constructului în termeni comportamentali concreţi. Operaţia se numeşte explicarea constructului şi constă din trei paşi (Murphy & Davidshofer, 1991):

1. Se identifică acele comportamente care au legătură cu constructul măsurat de test.
2. Se identifică alte constructe și pentru fiecare se decide dacă are sau nu legătură cu constructul măsurat de test.
3. Se alcătuiește câte o listă de comportamente prin care se manifestă aceste constructe. Pentru fiecare din ele, pe baza relațiilor dintre constructe, se decide dacă are sau nu legătură cu constructul măsurat de test.

Un exemplu. În Tabelul 3-5, Murpy și Davidshofer (1991) exemplifică procesul de descriere a constructului de "agresivitate la elevi".

Din analiza Tabelului 3-5 rezultă un sistem de relații care înglobează un set de constructe și comportamente, denumit de L. J. Cronbach, "rețea nomologică" ("nomologică" în sensul că ea are la bază declarații exprimate sub formă de legi). Unii autori au folosit termenul de "stohastic", susținând că relațiile dintre componentele rețelei sunt mai degrabă probabilistice decât de tip lege.

Ideile referitoare la caracterul structural al constructelor și la existența rețelelor nomologice au condus la descompunerea operației de validare a testului relativă la constructul măsurat, în trei componente :

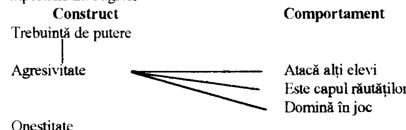
- validarea materialului testului (substantive validity), care necesită specificarea constructului măsurat de test și se confundă uneori cu validarea conținutului testului;
- validarea structurii constructului aflat la baza testului (structural validity), care identifică elementele componente ale constructului și investighează legăturile dintre ele;
- validarea externă (external validity), care urmărește să determine relațiile constructului cu alte măsurători (constructe sau variabile observate).

Tabelul 3-5
Etapele de descriere a a constructului de "agresivitate la elevi" (Murphy & Davidshofer, 1991)

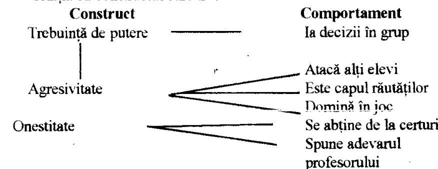
1. Identificarea comportamentelor referitoare la agresivitate



2. Identificarea altor constructe și se decide dacă au sau nu legătură cu comportamentul agresiv



3. Se identifică comportamentele aferente fiecărui construct și se determină relația cu constructul care trebuie măsurat.



Notă: constructele care au legătură unele cu altele sunt legate printr-un segment vertical.

Validarea externă se referă la calitatea testului de a avea relații corecte cu alte măsurători psihologice; ea este denumită uneori "validitate convergentă și discriminantă" (Murphy & Davidshofer, 1991).

Un test are validitate convergentă dacă măsoară ceea ce evaluează și alte teste sau variabile care se referă la același construct, deci scorurile sale variază (liniar sau nu) în același sens cu rezultatele respectivelor măsurători. Un test are validitate discriminantă dacă evaluează altceva decât diverse teste sau variabile despre care se știe că se referă la constructe ce nu au legătură cu constructul măsurat de test. Aceasta înseamnă că scorurile testului nu sunt în relație funcțională monotonă (liniară sau nu) cu rezultatele acelor măsurători.

Verificarea validității convergente și discriminante se realizează conform următorului algoritm:

- Se identifică acele comportamente care sunt legate de constructul măsurat de test.
- Se determină constructele care au legături cu constructul măsurat de test și se pun în evidență acele comportamente legate de ele care pot fi conectate cu constructul măsurat de test.
- Pentru toate comportamentele reținute se apreciază ce fel de corelații ar putea exista între măsurile lor și o măsură bună a constructului care ne interesează (corelații pozitive, negative, mari, mici sau nule).
- Se determină coeficienții de corelație între scorurile la testul studiat și rezultatele unor măsurători (eventual scoruri de teste) efectuate asupra comportamentelor reținute.
- Se compară coeficienții de corelație obținuți cu valorile așteptate.

O metodă utilizată pentru verificarea validității convergente și discriminante a testelor este aceea propusă de D. T. Campbell și D. W. Fiske, bazată pe matricea "multitrait-multimethod" (Anastasi, 1976; Pitariu, 1994). Aceasta conține valorile coeficienților de corelație liniară între diverse măsurători ale acelorași constructe dar

obținute prin intermediul a două sau trei tehnici ori instrumente diferite.

Exemplu

În urma evaluării constructelor A, B și C, despre care se face presupunerea că nu au legături între ele, prin trei metode diferite - "test I", "test II" și "test III"- s-a obținut următoarea matrice de coeficienți de corelație liniară (pentru a se putea urmări mai ușor datele, ea a fost scrisă sub forma unui tabel):

		Test I A	Test I B	Test I C	Test II A	Test II B	Test II C	Test III A	Test III B	Test III C
Test I	A	.81								
	B	.12	.86							
	C	.04	.32	.76						
Test II	A	.46			.74					
	B	.14	.48		.02	.77				
	C	.08	.21	.55	.10	.15	.82			
Test III	A	.50			.54			.85		
	B	.32	.46		.37	.65		.38	.83	
	C	.28	.36	.60	.28	.38	.61	.36	.40	.85

Coeficienții cu valori mai mari decât .35 sunt semnificativi la pragul p=.05.

Valorile notate pe diagonala principală sunt coeficienții de fidelitate ai testelor, obținuți prin metoda formelor paralele. De exemplu, pentru testul II, coeficienții de fidelitate sunt .74 (pentru constructul A), .77 (pentru constructul B) și .82 (pentru constructul C).Valorile de pe diagonalele celorlalte căsuțe sunt coeficienții de corelație între măsurătorile realizate de cele trei teste pentru aceleași constructe. Ei servesc la verificarea validității convergente. De exemplu, pentru constructul A, coeficienții de corelație sunt:

.46 (testul I cu testul II);

.50 (testul I cu testul III);
.54 (testul II cu testul III).

Toţi aceşti coeficienţi sunt semnificativi statistic la pragul p=.05, ceea ce denotă o bună validitate convergentă a celor trei teste.

Celelalte valori din matrice reprezintă coeficienţii de corelaţie liniară între constructe diferite, măsurate prin acelaşi test sau prin teste diferite. Ei servesc la aprecierea validităţii discriminante a testelor.

Se observă că testele I şi II discriminează între constructele A şi B (coeficienţii de corelaţie aflaţi sub diagonala căsuţelor I-I, I-II şi II-II sunt nesemnificativi la pragul p=.05), în timp ce testul III nu are validitate discriminantă (coeficienţii de corelaţie liniară între constructele A, B şi C, măsurate prin testul III -în căsuţa III-III- sau prin testul III şi un alt test - în căsuţele I-III şi II-III- sunt, în general, semnificativi la pragul p=.05).

În exemplul prezentat, validitatea convergentă şi discriminantă a testelor I, II şi III a fost uşor de dedus. Dar, în practică, apar situaţii complexe, când mulţi coeficienţi de corelaţie liniară pentru un acelaşi construct sunt nesemnificativi şi mulţi coeficienţi de corelaţie liniară între constructe diferite sunt semnificativi statistic. În asemenea cazuri este greu de apreciat care dintre teste este lipsit de validitate.

Dacă testele măsoară constructe cu o structură complexă, pot să apară dificultăţi la interpretarea scorurilor. Din acest motiv, unii autori, ca M. L. Tenopyr, D. W. Fiske, J. C. Nunnally şi R. L. Durham, susţin că pentru a putea utiliza la interpretarea scorurilor unui test deducţii bazate pe felul în care este construit testul, trebuie ca testul să se refere la constructe simple, bine definite şi cu manifestări uşor de observat (Silva, 1993).

Din cele prezentate mai sus se constată că întregul proces de validare relativă la construct a testului constă în formularea de ipoteze

ştiinţifice (asupra componentelor constructului, asupra relaţiilor dintre aceste componente etc.) şi în verificarea acestora. Metodele aplicate în validare sunt dependente de ipotezele formulate. Din acest motiv nu există metode valabile pentru validarea relativă la construct a oricărui test. De exemplu, validarea unui test relativă la constructul măsurat poate să conţină următorii paşi (Fisseni, 1990):

- se identifică la care aspect din teorie se referă fiecare item;
- se verifică dacă itemii sunt formulaţi clar;
- se studiază corelaţiile dintre itemi;
- se cercetează corelaţiile itemilor cu o scală externă testului;
- se face analiza de itemi;
- se efectuează analiza factorială şi analiza de clusteri a itemilor.

O altă metodă de studiere a validităţii relative la construct implică manipularea experimentală a constructului măsurat de test (Murphy, 1987). De exemplu, în cazul unui test destinat să măsoare anxietatea, se formează două grupe similare de subiecţi. În condiţii care nu generează anxietate, cele două grupe vor realiza la test scoruri asemănătoare. Dacă, însă la administrarea testului, unui grup i se spune că va avea o surpriză neplăcută, iar celuilalt i se face instructajul obişnuit, este de aşteptat ca scorurile testului să difere semnificativ între grupe.

Trebuie să se reţină că validitatea relativă la construct a unui test nu poate fi exprimată printr-o mulţime de coeficienţi. Ea se deduce prin acumularea rezultatelor cercetării relative la ipotezele propuse (Silva, 1993).

Validitatea testului relativă la conţinutul său

Acest tip de validitate este analizat la testele care se utilizează pentru a estima "cum acţionează o persoană în universul de situaţii pe care testul intenţionează să îl reprezinte" (APA Standards, 1974).

În literatura psihologică există păreri diferite despre ceea ce se urmăreşe prin validarea testului relativă la conţinut. Anastasi (1976), consideră că acest tip de validare implică "examinarea sistematică a conţinutului testului pentru a determina dacă el acoperă un eşantion reprezentativ din domeniul de comportamente pe care testul în cauză trebuie să îl măsoare". Alţi autori înlocuiesc expresia "domeniu de comportamente" cu "mulţimea sarcinilor" sau "mulţimea performanţelor". M. M. Linehan susţine că "validarea relativă la conţinut determină cât de bine reprezintă condiţiile în care este observat (prin test) comportamentul persoanei toate mulţimile de condiţii pe care vrem să le generalizăm" (Silva, 1993). R. T. Lennon consideră că validitatea relativă la conţinut se referă la răspunsurile subiectului şi nu la întrebările testului. Pentru analizarea acestui tip de validitate trebuie luat în considerare nu numai conţinutul itemilor testului ci şi procesul utilizat de subiect pentru a ajunge la formularea răpunsului (Silva, 1993). De exemplu, un test având drept domeniu de conţinut "cunoştinţele de ortografie", format din itemi cu alegere multiplă, poate verifica dacă un individ recunoaşte cuvintele scrise corect din punct de vedere ortografic, dar nu şi dacă acesta ştie să scrie corect după dictare. Deci nu este suficient să se spună că testul evaluează cunoştinţele de ortografie.

În unele accepţiuni, validarea relativă la conţinut este o componentă a validării relative la construct. Ea se realizează după definirea constructului şi urmăreşte să verifice dacă eşantionul de stimuli şi cel de răspunsuri observate şi înregistrate în procesul de măsurare sunt reprezentative pentru universul de comportamente pe care îl defineşte constructul respectiv (M. M. Linehan, după Silva, 1993).

Adoptând punctul de vedere formulat de Anastasi (1976), analiza validităţii relative la conţinut a unui test impune ca prim pas definirea şi descrierea domeniului de conţinut al testului. Este esenţial ca descrierea să fie cât mai detaliată şi să fie precizate clar limitele acestuia.

Pasul următor îl constituie analizarea itemilor testului. Sunt eliminaţi acei itemi care nu se referă la comportamente cuprinse în domeniul de conţinut. Pentru itemii rămaşi se identifică zonele domeniului de conţinut pe care le măsoară.

La această operaţie este bine să participe cât mai mulţi experţi, care să aprecieze pentru fiecare item dacă acesta se referă la domeniul de conţinut al testului şi dacă este relevant. O analiză atentă o pretind itemii complecşi; ei pot să disperseze atenţia subiecţilor şi, prin urmare, scorul lor să fie contaminat de alte caracteristici decât aceea care ar trebui să fie măsurată.

Exemplu

Li se poate cere elevilor să efectueze aceleaşi operaţii aritmetice formulând două probleme diferite:

a. 3+5-2-4=?

b. Vasile are 3 mere. Corina îi dă 5 portocale în schimbul a 2 mere. După ce mănâncă 4 portocale, câte fructe mai are Vasile ?

Cea de a doua problemă le propune subiecţilor o sarcină mai complexă decât prima, întrucât le pretinde să citească un text mai lung. Dacă ea este inclusă într-un test de raţionament aritmetic pentru elevi, este posibil ca scorul său să fie alterat de variabila irelevantă "deprinderea de a citi".

Din acest motiv este important să se verifice pentru fiecare item dacă nu cumva necesită un comportament care nu aparţine domeniului de conţinut al testului.

Ultima etapă a validării constă în compararea structurii testului cu structura domeniului de conținut. Se verifică dacă sunt acoperite toate zonele domeniului de conținut și dacă itemii sunt repartizați proporțional cu mărimea și importanța zonelor.

Aprecierea generală privind validitatea testului relativă la conținutul său este subiectivă. Ea nu apelează la metode statistice, ci se bazează doar pe raționamente. Din acest motiv, validarea relativă la conținut este denumită în literatură și "validare logică" sau "validare rațională" (Thorndike & Hagen, 1961; Anastasi, 1976).

Totuși, în cazul testelor care verifică deprinderi sau cunoștințe însușite în cadrul unui program de instruire există recomandarea de a lua în considerare, la aprecierea validității relative la conținut a testului, și rezultatul comparării statistice a scorurilor obținute la test de aceleași persoane, înainte și după acest program (Anastasi, 1954). Într-adevăr, dacă există o suprapunere între conținutul itemilor testului și conținutul programului de instruire și dacă itemii sunt bine construiți, atunci este de așteptat ca scorurile finale să fie mult mai mari decât cele inițiale, deci media diferențelor de scor să fie pozitivă și să difere semnificativ de zero.

Validarea relativă la conținut se impune în cazul testelor utilizate la măsurarea unui atribut ce nu poate fi exprimat printr-un construct. De exemplu, putem descrie destul de detaliat domeniul de conținut pentru "cunoștințe de statistică descriptivă", dar este foarte dificil să găsim comportamentele prin care această caracteristică să poată fi observată (Murphy & Davidshofer, 1991).

Trebuie menționat că validarea relativă la conținut nu permite să se tragă nici o concluzie despre calitatea măsurătorii realizate de test (Murphy & Davidshofer, 1991). De asemenea, ea nu furnizează suficiente informații pentru a putea interpreta corect scorurile testului.

Nu se poate afirma că, în privința atributului măsurat de test, un subiect care a realizat un scor mai mare îi este superior unuia care a obținut un scor mai mic.

Deci, operațiile cuprinse sub denumirea de "validarea testului relativă la conținutul său" nu reprezintă o validare, în sensul definiției date de Standardele APA din 1985. Ele realizează doar analiza testului în privința clarității, a reprezentativității și a relevanței conținutului său. Ar putea fi privite ca o componentă a validării testului.

Validitatea testului relativă la criteriu

Procedee utilizate pentru validare. Validarea relativă la criteriu se referă la gradul în care deducțiile făcute pornind de la scorurile testului concordă cu cele bazate pe scorurile unei alte măsurări, numită criteriu.

Este important ca un test să fie valid relativ la un criteriu atunci când testul trebuie să înlocuiască variabila criteriu, fie din cauză că valorile variabilei criteriu nu sunt cunoscute în momentul testării (de exemplu, în cadrul unei acțiuni de selecție profesională), fie pentru că administrarea testului este mai puțin costisitoare decât măsurarea variabilei criteriu.

Dacă variabila criteriu este un alt test psihologic, este necesar ca acesta să fie fidel și, în plus, pentru situația în care are loc testarea să existe siguranța că scorurile sale permit o interpretare corectă (deci criteriul este valid) (Fisseni, 1990).

Aprecierea validității relative la criteriu se bazează, de obicei, pe valorile unui "coeficient de validitate a testului", obținut în urma calculului coeficientului de corelație (liniară sau de alt fel) între scorurilor testului și rezultatele criteriului.

În cazul testelor utilizate pentru măsurare, validarea relativă la criteriu este o componentă a validării relative la constructul măsurat. Ea apreciază "gradul în care rezultatele furnizate de test corelează cu rezultatele altui instrument de evaluare despre care se presupune că măsoară același construct sau unul similar" (S. N. Haynes, citat după Silva, 1993). Evident, calculul coeficientului de corelație trebuie să fie precedat de analiza constructelor măsurate de test și de criteriu.

Chiar și "validarea factorială" utilizată pentru cercetarea validității relative la construct a unui test este, de fapt, o validare relativă la criteriu (Anastasi, 1954). Criteriul este reprezentat, în acest caz, de factorul care corespunde constructului măsurat de test. Una dintre condițiile impuse pentru a putea aprecia că testul este valid este ca itemii săi să aibă saturații mari în factorul respectiv. Or, saturația unui item într-un factor este valoarea coeficientului de corelație liniară între scorurile itemului și scorurile factorului.

În cazul testelor utilizate pentru a se lua decizii, validarea relativă la criteriu urmărește dacă deciziile bazate pe scorurile testului coincid cu cele care au la bază valorile unei variabile criteriu, despre care se presupune că permite să se ia decizii corecte. De exemplu, decizia poate consta în acceptarea sau respingerea candidaților care concurează pentru un post. Se dovedește că testul este valid relativ la acest criteriu dacă se constată că indivizii acceptați (pe baza unor considerente care nu includ rezultatele testului) sunt tocmai cei care au scoruri mari la test, iar cei respinși sunt cei cu scoruri mici.

Variabila criteriu poate fi unidimensională sau multidimensională, cu valori cantitative sau/și calitative. De exemplu, în selecția profesională o variabilă criteriu poate fi constituită din performanțele în activitatea anterioară (apreciate printr-un calificativ), rezultatele unor probe de lucru (cu valori cantitative) și unele date biografice (apreciate dihotomic, după cum îndeplinesc sau nu anumite condiții fixate anterior; de exemplu sexul, starea civilă etc.).

Validarea testelor utilizate pentru predicție are în vedere gradul în care scorurile testului -numit în acest caz "predictor"- pot fi utilizate pentru a se deduce performanța unei variabile numită "criteriu", independentă de test (Guion, 1974). Trebuie menționat că o aceeași variabilă poate să joace rol de predictor într-un caz (de exemplu, performanța școlară ca predictor pentru performanța profesională) sau rol de criteriu în alt caz (de exemplu, aceeași performanță școlară ca și criteriu pentru un test de inteligență).

Există două strategii utilizate în mod obișnuit pentru aprecierea validității relative la criteriu a testelor:

- validarea predictivă;
- validarea concurentă.

Validarea predictivă este interesată de gradul de eficiență al testului în prognoza comportamentului unei persoane într-o situație dată. Ea se realizează prin compararea scorurilor testului cu rezultatele obținute la criteriu după un interval de timp mai lung sau mai scurt.

De exemplu, pentru estimarea validității predictive a unui test utilizat în selecția profesională se calculează coeficientul de corelație liniară între scorurile testului, administrat, de pildă, cu ocazia finalizării unui curs de pregătire profesională, și măsurile performanțelor obținute de aceleași persoane după un anumit timp, când tuturora li s-a creat posibilitatea să aplice cunoștințele respective.
Trebuie subliniat faptul că validitatea predictivă se cercetează folosind loturi neselecționate. Aceasta înseamnă că toate persoanele cărora li s-a administrat testul vor urma același tratament și vor fi prezente la evaluarea variabilei criteriu. În acest fel se asigură ca în lotul de subiecți care servește la compararea testului cu criteriul să existe atât scoruri mari la test, cât și scoruri mici.
Nu totdeauna se poate realiza așa ceva. Este greu de acceptat, de exemplu, ca să fie angajate toate persoanele care se prezintă la

selecția profesională sau ca decizia referitoare la angajare să se ia la întâmplare.
Pentru estimarea validității concurente este necesară obținerea scorurilor la criteriu aproximativ în același timp cu cele ale testului.

Dar, deosebirea esențială dintre cele două tipuri de validări relative la criteriu nu constă în lungimea intervalului de timp dintre administrarea testului și înregistrarea rezultatelor criteriului, ci în faptul că validarea concurentă utilizează un eșantion de populație selecționat (personal angajat deja sau studenți admiși în facultăți). Apare aici problema denumită "restrângerea mulțimii", pe care o vom explica în continuare, și care face ca validarea concurentă să nu poată furniza totdeauna informații corecte asupra capacității predictive a testului.

Problema restrângerii mulțimii. Lotul de persoane utilizat la calculul coeficientului de corelație liniară provine din populația asupra căreia ar fi trebuit să se facă predicția, dar până în momentul când se înregistrează valorile variabilei criteriu el a fost supus deja la două selecții:

- Pe de o parte, au fost reținuți din eșantionul inițial doar cei care au obținut rezultate bune la un anumit criteriu (de exemplu, cei care au avut note mari la admiterea în facultate). Dacă acest criteriu măsoară același lucru ca și testul, atunci persoanele rămase sunt cele care vor obține la test scoruri mai mari decât ar fi realizat cei respinși.
- Pe de altă parte, au fost eliminați cei cu performanțe școlare sau profesionale slabe, adică cei cu valori mici la variabila criteriu utilizată în studiul validității. Este posibil să fi fost înlăturați prin promovare, și cei cu valori foarte mari la criteriu.

Deci, pentru lotul considerat, atât mulțimea scorurilor testului cât și mulțimea valorilor criteriului reprezintă restrângeri ale mulțimilor

corespunzătoare care s-ar fi obținut pentru întreaga populație. În consecință, coeficientul de corelație liniară calculat pe baza persoanelor selecționate diferă de cel care ar fi rezultat lucrând cu întreaga populație și care ar fi fost util pentru aprecierea valorii predictive a testului.

O asemenea situație este prezentată în Figura 3-5 (Murphy & Davidshofer 1991).

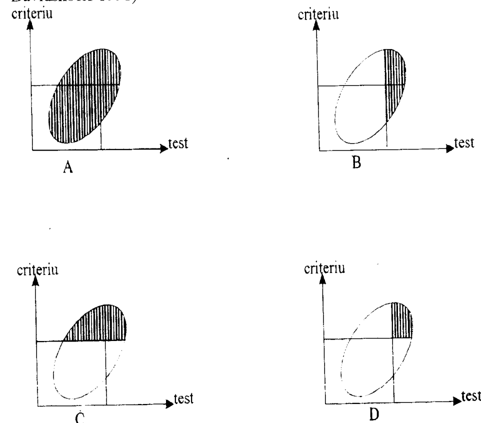


Figura 3-5. Influența "restrângerii mulțimii" asupra formei norului statistic și a mărimii coeficientului de corelație.

A. În populația P scorurile testului corelează destul de puternic cu valorile variabilei criteriu (r=0.75).
B. Dacă din populația P este selectat grupul celor cu scoruri mari la test, mulțimea valorilor predictorului se restrânge și corelația scade (r=0.28).
C. Restrângerea mulțimii valorilor variabilei criteriu are efecte similare (r=0.30).
D. Dacă se selecționează persoanele și în funcție de predictor și în funcție de criteriu, corelația se apropie de zero (r=0.09).

Problema restrângerii mulțimii i-a făcut pe unii specialiști în teoria testelor să recomande utilizarea validării predictive pentru testele care servesc la prognoză și a validării concurente pentru testele aplicate în scop de diagnoză. De exemplu, se va efectua validarea concurentă dacă testul trebuie să ne permită să răspundem la întrebarea "Este Ion neurotic ?" și se va realiza validarea predictivă, dacă ne interesează să răspundem la întrebarea "Este de așteptat ca Ion să devină neurotic?" (Murphy & Davidshofer 1991).

Statistici utilizate în validarea relativă la criteriu a unui test psihologic

Coeficientul de validitate al unui test, conform definiției din Dictionary of Education (1959) și din majoritatea lucrărilor de teoria testelor, este egal cu coeficientul de corelație liniară între scorurile testului și rezultatele unei măsurători considerate drept variabilă criteriu. Unii autori (Lord & Novick, 1968) înțeleg prin coeficient de validitate valoarea absolută a coeficientului de corelație liniară; alți autori (Cureton, 1955) impun calculul coeficientului de corelație între scorurile observate la test și scorurile reale ale variabilei criteriu.

În literatura de specialitate există și recomandarea de a calcula coeficientul de validitate folosind alți indici în locul coeficientului de corelație liniară. Anastasi (1976) propune ca atunci când testul și criteriul sunt variabile continue, validitatea să se calculeze prin formula coeficientului de corelație liniară, iar în celelalte cazuri, să se folosească alte tipuri de coeficienți. Cronbach (1966) permite calcularea validității relative la criteriu cu ajutorul coeficientului de corelație a rangurilor; de remarcat că, dacă regresia dintre test și criteriu este o funcție monotonă crescătoare, coeficientul de corelație a rangurilor are o valoare apropiată de 1, deci testul este valid în sensul lui Cronbach.
Se apreciază că un test are validitate bună atunci când coeficientul de validitate este mare.

Unii autori (Hammond, 1995) recomandă ca validarea unui test relativă la criteriu să utilizeze mai multe variabile criteriu, dintre care unele nu au legătură cu ceea ce testul măsoară sau prezice. Acestea servesc la verificarea specificității testului. Este de dorit ca testul să nu coreleze cu ele.

O altă posibilitate de cercetare a validității unui test relativă la criteriu utilizează două grupe contrastante de subiecți, formate pe baza valorilor criteriului: una este compusă din indivizi cu rezultate foarte bune, cealaltă din persoane cu rezultate foarte slabe. Se compară statistic mediile scorurilor obținute la test de persoanele din cele două grupe și se apreciază că testul este valid relativ la criteriu dacă mediile diferă semnificativ, la un prag de semnificație fixat (Anastasi, 1976). Această metodă se aplică mai ales atunci când variabila criteriu este calitativă - deci nu se poate calcula coeficientul de corelație liniară - și atunci când criteriul este compus din mai multe variabile, cantitative și/sau calitative (de exemplu, într-o selecție profesională criteriul poate consta din câteva teste de aptitudini, un interviu și referințe de la ultimul loc de muncă).

Interpretarea coeficienților de validitate. În experiența obișnuită, valoarea coeficientului de validitate predictivă al unui test este cuprinsă între 0 și 0.60, în majoritatea cazurilor fiind situată în partea inferioară a intervalului (0, 0.60] (Guilford, 1965).

Se pot obține valori mari ale coeficienților de validitate atunci când testul predictor reproduce o situație din criteriu (de exemplu, dacă se dă un test de aptitudini matematice unor studenți de la Politehnică). Din contră, atunci când testul se referă la atribute dificil de evaluat prin orice fel de date non-test (de exemplu, adaptarea socială și emoțională, creativitatea etc.) coeficienții de validitate iau valori mai mici decât 0.30 (L. J. Cronbach, C. Gleser, după Silva, 1993).

C.L. Hull (Guilford, 1965) a impus condiția ca un test să fie utilizat în practică doar dacă are un coeficient de validitate mai mare decât 0.45; această cerință este însă greu de realizat. Un test bun rareori poate da, cu un criteriu important, un coeficient de corelație mai mare decât 0.5 (Murphy & Davidshofer, 1991).

Un test ideal din punctul de vedere al validității este cel ai cărui itemi măsoară, fiecare, un alt factor; deci este un test cu consistență internă scăzută.

CAPITOLUL 4

ANALIZA DE ITEMI

Construcția testelor, indiferent dacă sunt teste de cunoștințe sau teste psihologice pentru măsurarea aptitudinilor ori chestionare de personalitate, presupune organizarea unor activități specifice, destinate elaborării cît mai adecvate a acestora. Este vorba de activități de mare complexitate, care sunt foarte importante pentru că asigură calitatea instrumentului diagnostic respectiv. Analiza de itemi este una dintre aceste activități. O dată cu facilitățile aduse de calculator pentru efectuarea prelucrărilor statistice, ea a cunoscut o expansiune impresionantă, fiind dezvoltate tehnici de mare sensibilitate pentru evaluarea itemilor care intră în componența unui test. Teoria răspunsului la itemi (TRI) împreună cu alte aplicații de analiză calitativă și cantitativă în proiectarea testelor sunt abordări mai noi, de mare eficacitate. În ceea ce urmează ne vom opri asupra analizei de itemi, ca o etapă necesară a construcției unui instrument psihodiagnostic.

Discutînd despre "analiza de itemi", ne referim la un set structurat de prelucrări statistice care se aplică la fiecare item al unui test. Alegerea procedurilor statistice și interpretarea rezultatelor obținute sunt determinate de obiectivul testării și de tipul itemilor care intră în componența testului respectiv. Bineînțeles, există cîteva proceduri generale pe care le vom regăsi în orice analiză de itemi.

Analiza de itemi se ocupă în mod esențial de descifrarea mecanismului funcționării itemilor testului, a semnificației lor, a calităților acestora ca instrumente de măsură și a capacității lor predictive. Analiza de itemi furnizează informațiile pe baza cărora

se aleg itemii care intră în componenţa testului. Ea este o etapă necesară şi deosebit de importantă pentru construirea unui test de bună calitate.

Cînd se procedează la construirea unui test, proiectantul este pus în situaţia să răspundă la cîteva întrebări : "Care este mecanismul de alegere/formulare a răspunsului ?", "Cîţi din cei testaţi vor răspunde corect la item ?", "Cît de bine măsoară itemul ceea ce îşi propune să măsoare ?" sau "Există deosebiri între modul cum răspund la item cei care obţin scoruri mari la întregul test şi cei care realizează scoruri mici ?" Concret, pentru a răspunde la toate aceste întrebări au fost dezvoltate proceduri statistice şi tehnici specifice de investigare: analiza răspunsurilor incorecte, studiul indicelui de dificultate al itemilor, al validităţii itemilor şi al puterii lor de discriminare (pentru o tratare aprofundată a acestei probleme a se consulta Albu şi Pitariu, 1993).

Menţionăm că la baza proiectării unui instrument psihodiagnostic stă întotdeauna o teorie, un cadru teoretic pe care se fundamentează atît construcţia lui, cît şi interpretarea rezultatelor acestuia. Adesea, asemenea teorii au determinat elaborarea şi dezvoltarea unor proceduri statistice, care servesc la construirea testelor sau a chestionarelor de personalitate. De exemplu, teoria trăsăturilor a stat la baza dezvoltării analizei factoriale, teoria stare-trăsătură a dus la dezvoltarea şi utilizarea pe scară largă a modelelor cauzale etc.

Activitatea de proiectare a testelor psihologice presupune parcurgerea unei succesiuni de secvenţe aproximativ comune, dar cu conţinut adesea diferit. Astfel, orice debut în construcţia unui test înseamnă racordarea acestuia la o teorie psihologică şi întrunirea unui grup de experţi care să se implice în formularea itemilor testului respectiv (munca în grup este recomandată datorită productivităţii ei mai mari). După ce întrebările (sau

itemii) care urmează să facă parte din test au fost analizate şi criticate de un alt grup de experţi, apoi revizuite pe baza observaţiilor formulate de acesta, este necesar ca ele să fie experimentate pe un eşantion de persoane, asemănător populaţiei căreia îi este destinată forma finală a testului.

Rezultatele care se vor obţine vor face obiectul unei analize de itemi, urmărindu-se (Albu & Pitariu, 1993):

- analiza frecvenţei cu care au fost alese răspunsurile, studierea cauzelor alegerilor preferenţiale şi a respingerilor, făcîndu-se totodată înlocuirile necesare;
- identificarea întrebărilor/itemilor al căror text conţine termeni ambigui, care propun răspunsuri neplauzibile (pe care nu le alege nimeni) sau evidente (pe care le aleg toţi cei testaţi) şi a celor care nu oferă suficiente informaţii pentru a se putea stabili răspunsul corect;
- determinarea validităţii fiecărei întrebări/item, pentru a păstra în testul final doar itemii care măsoară ceea ce se doreşte;
- determinarea indicelui de dificultate pentru fiecare întrebare/item, spre a putea selecta itemii care, în ceea ce priveşte dificultatea, corespund scopului testului final;
- determinarea puterii de discriminare a fiecărui item, pentru a-i reţine doar pe cei care pot face deosebirea între examinaţii care dispun şi cei care nu dispun de cunoştinţele, calităţile psihice sau deprinderile verificate de test.

Itemii care se dovedesc nerelevanți după analiza de itemi efectuată vor fi eliminați. Itemii reținuți vor fi introduși într-o bancă de itemi, care stă la baza construcției testului psihologic sau de cunoștințe sau a chestionarului de personalitate; adesea este nevoie de ei pentru construirea unor forme paralele.

Analiza de itemi, așa cum am mai arătat, ne ajută să înțelegem mai bine testul, fidelitatea și validitatea sa. Prin studiul fidelității se verifică existența unor surse de erori de măsurare cu efecte asupra rezultatelor testului. Prin cercetarea validității se urmărește dacă scorurile testului permit o interpretare corectă în scop de predicție sau de măsurare. Ori de cîte ori se constată distorsiuni la nivelul fidelității sau al validității, din analiza de itemi se pot obține informații asupra cauzelor acestora și a posibilităților de îmbunătățire a testului.

Este important de reținut, deci, că analiza de itemi ne poate sugera o serie de mijloace de localizare a defectelor constructive ale testului, indicîndu-ne posibilitățile de optimizare a acestuia. Prin tehnicile utilizate de analiza de itemi putem elimina itemii necorespunzători și să-i păstrăm numai pe cei relevanți. De exemplu, teoria fidelității susține că toți itemii unui test sau ai unei scale a unui chestionar de personalitate trebuie să măsoare același lucru (memoria mecanică, raționamentul verbal, dominanța, anxietatea etc.). Prin analiza de itemi avem posibilitatea să identificăm itemii care distorsionează scorul probei respective și să-i eliminăm.

Analiza de itemi este o sursă de informații bogată despre validitatea testului relativă la constructul măsurat. Carroll (1979) compară itemi de diferite dificultăți și urmărește felul în care itemii mai dificili pretind o augmentare a abilităților de procesare mentală sau de rezolvare de probleme. Psihologii cognitiviști au

studiat problema procesării informațiilor cu ocazia investigării patternelor de performanță la sarcinile cognitive implicate în rezolvarea itemilor de test, cu varierea sistematică a unor atribute (Reșnick, 1976; Sternberg, 1979). Folosind metode ca analiza de conținut a protocoalelor verbale sau scrise, modelarea matematică a timpului de răspuns sau a erorilor și simularea pe calculator a proceselor cognitive, ei au formulat o serie de inferențe despre natura abilităților mintale și nivelul de procesare a informației pe parcursul completării unui test.

Particularități ale itemilor unui test

Examinarea oricărui item a unui test trebuie să aibă în vedere *ce măsoară* acesta, *cum măsoară* și ce informații putem obține din modul în care s-a răspuns la item. Analiza acestor aspecte se bazează pe patru tipuri de măsurări:

- Măsurări ale distragerii atenției (Cîte persoane aleg fiecare dintre răspunsurile propuse, la itemii cu răspunsuri la alegere? - de obicei unul dintre răspunsuri este corect, celelalte fiind așa-numitele *răspunsuri distorsionante*).
- Măsurări ale dificultății (Cîte persoane răspund corect la item? - discutăm în acest caz despre *indicele de dificultate al itemului*).
- Măsurări ale validității itemului (Itemii testului măsoară ceea ce trebuie să măsoare? - în acest caz se discută despre *validitatea itemului*).
- Măsurări ale capacității de discriminare a itemului (Două grupe de subiecți, contrastante în privința a ceea ce testul trebuie să măsoare, se deosebesc între ele prin modul de a răspunde la item? - discutăm în acest caz despre *indicele de discriminare al itemului*).

Analiza itemilor distorsionanți sau a variantelor incorecte de răspuns

Testele de cunoștințe sau de aptitudini și chestionarele de personalitate sunt alcătuite frecvent din itemi cu alegere multiplă (multiple-choice item). Aceștia au forma unei întrebări la care sunt atașate mai multe răspunsuri, dintre care numai unul este cel corect sau potrivit. Frecvența cu care subiecții aleg o variantă sau alta de răspuns este o sursă de informații deosebit de utilă despre calitatea construcției itemului respectiv.

Să analizăm un exemplu:

Itemul 20. Care dintre cifrele de mai jos completează logic raționamentul:

RAARRAR:3113313 = RARAARAR:?

(a) 13133131 (d) 31311313
(b) 13133313 (e) 31313113
(c) 31311131

Iată răspunsurile date de 150 subiecți (Tabelul 4-1):

Tabelul 4-1

Repartiția răspunsurilor la itemul unui test de aptitudini

Răspuns	Frecvența cu care s-a ales fiecare răspuns	Procentul alegerii fiecărui răspuns
a	15	10
b	10	7
c	22	15
Răspunsul corect d	75	50
e	28	19

Examinarea tabelului de mai sus ne indică faptul că 50% dintre subiecți au dat răspunsul corect (*d*). Un număr mai mic au ales varianta de răspuns *c* sau *e* și foarte puțini *a* sau *b*.

Un item construit în mod ideal propune variante de răspuns care sunt, toate, la fel de plauzibile pentru orice persoană care nu cunoaște răspunsul corect. Prin urmare, subiecții care cunosc răspunsul corect îl vor alege, iar ceilalți vor răspunde la întîmplare prin alegerea oricăreia dintre variantele de răspuns oferite. Aceasta înseamnă, pe de o parte, că unii subiecți vor ghici răspunsul corect, iar pe de altă parte, că fiecare din variantele incorecte de răspuns va fi aleasă cu aceeași frecvență, toate fiind la fel de populare. Referindu-ne la exemplul nostru, 50 % dintre persoanele examinate au
dat un răspuns corect și 50% au ales varianta *a, b, c* sau *e*. Dacă itemul ar fi unul construit ideal, frecvențele răspunsurilor eronate ar fi egale, fiecare din cele 4 răspunsuri propuse fiind ales de aproximativ 13% dintre subiecți.

Studiul variantelor eronate de răspuns sau a răspunsurilor distorsionante ne ajută să explicăm atît tipul de erori pe care le poate comite un subiect, cît și strategia sa de gîndire sau de rezolvare a problemei. În cazul testelor de cunoștințe selecția răspunsurilor ne informează asupra lacunelor în învățare. Pentru a efectua o analiză de acest fel, calculăm frecvența persoanelor care ar fi ales fiecare dintre variantele eronate de răspuns, dacă toate acestea i-ar fi părut la fel de atractive unui subiect care nu cunoaște răspunsul corect. Această frecvență (*F*) este egală cu raportul dintre numărul persoanelor care au răspuns greșit la item (*G*) și numărul total de răspunsuri eronate propuse de item (*e*). Formula de calcul este, deci, următoarea:

$$F = G / e$$

În cazul exemplului nostru, 75 de subiecți au dat răspunsuri incorecte. Practic ne așteptam ca fiecare dintre cele patru răspunsuri distorsionante să fi fost ales de 19 persoane (75:4). Dar, 22 de subiecți au ales varianta de răspuns *c*, mult mai puțini variantele *a* și *b* și cei mai mulți varianta *e*. Dacă numărul celor care aleg o variantă de răspuns eronată este foarte mic, înseamnă că aceasta se deosebește mult de răspunsul corect, este "evident greșită" pentru majoritatea subiecților și trebuie înlocuită. Cînd numărul celor care aleg o anumită variantă de răspuns eronată este foarte mare, există două posibilități: fie că această variantă este apropiată ca semnificație de răspunsul corect (în situația itemului nostru, ea implică aceleași mecanisme de gîndire ca varianta corectă), nepotrivirea fiind neesențială, fie că itemul este incorect construit (varianta de răspuns considerată corectă este greșită sau varianta eronată analizată este, și ea, corectă). Orice abatere de la frecvența F, în plus sau în minus, constituie, în general, deficiențe constructive care trebuiesc remediate, întrucît ele afectează validitatea itemului și, în final, a testului în care acesta este inclus.

Indicele de dificultate al unui item

Dificultatea este interpretată adesea în termeni de complexitate sau de apel la cunoștințe specializate.

Să analizăm doi itemi dintr-un test de cunostinte:

1. (5 x 4) + 5 = ...

2. Dacă $(x + 5)^2 = (y + 1)^3$, care din următoarele soluții este corectă?
   I. Cînd x este 3, y = x.
   II. Dacă x este 1, y = 5.
   III. [illegible]

$x = (y+1)^{3/2} - 5$

(a) numai I; (b) numai II; (c) numai III; (d) I și II; (e) I și III.

(Brownstein & Weiner, 1955)

Pentru mulți dintre noi, itemul al doilea va părea ca fiind mai complex, deci mai dificil, deoarece presupune cunoștințe matematice mai bogate. Matematicianului, însă, problema i se pare simplă.

Să luăm un alt exemplu:

1. Cine a fost M. Eminescu?
2. Cine a fost W. Wundt?

Și în acest caz, mulți consideră itemul al doilea mai dificil decît primul, deoarece solicită cunoștințe mai specializate. Pentru un student la psihologie răspunsul la itemul al doilea nu constituie deloc o problemă dificilă, în schimb pentru un student de la matematică, da.

Din exemplele de mai sus rezultă că dificultatea unui item trebuie analizată avînd în vedere populația căreia îi este destinat testul din care face parte itemul. Adesea, ea este determinată de complexitatea sarcinii impuse de item și/sau de cunoștințele de strictă specialitate pe care acesta le pretinde.

În practica psihologică a proiectării de teste, în mod firesc, unii itemi sunt mai greu de rezolvat decît alții, dar este foarte greu să definim dificultatea în termeni de caracteristici intrinseci ale itemilor. Cea mai obișnuită măsură a dificultății unui item este frecvența relativă a persoanelor care au răspuns corect, din totalul persoanelor care au răspuns la itemul respectiv (Murphy & Davidshofer (1991). Indicele de dificultate astfel definit se notează cu *p*. *p* reprezintă scorul mediu care s-ar obține la item, în

lotul de persoane examinate cu el, dacă s-ar acorda un punct pentru fiecare răspuns corect și zero puncte pentru fiecare răspuns greșit sau omis. Se observă că p este cuprins întotdeauna între 0 și 1.

Formula de calcul a indicelui de dificultate al unui item din test este următoarea:

$$p = C / N$$

unde:
p = indicele de dificultate al itemului;
C = numărul persoanelor care au răspuns corect la item;
N = numărul persoanelor examinate cu testul din care face parte itemul.

Indicele de dificultate este o caracteristică atît a itemului, cît și a populației căreia i s-a administrat testul. Avem deci de-a face cu o dificultate fizică, datorată complexității sarcinii și una culturală; legată de domeniul de cunoștințe la care face apel itemul respectiv (de exemplu, la itemi cu conținut geografic vor răspunde mai bine geografii, la cei de matematică, matematicienii etc.). Din aceste considerente putem afirma că valoarea p este o măsură comportamentală (adică, un răspuns măsurabil al organismului).

Majoritatea testelor sunt compuse din itemi mai grei și mai ușori. Performanțele la test ale unei persoane sunt judecate, în general, prin scorul realizat, fără a se ține seama care sunt itemii la care aceasta a răspuns corect. Un item la care toată lumea răspunde corect nu produce nici o diferențiere între subiecți; varianța scorurilor sale este egală cu zero. Deci el este inutil. Același lucru este valabil și pentru un item la care nu poate răspunde nici un subiect. Itemii cu valoare lui p egală cu 0 (care nu au fost rezolvați corect de nici o persoană) sau cu 1 (rezolvați corect de toți cei examinați) afectează

media performanței lotului de subiecți, dar nu au nici un efect asupra validității sau a fidelității testului. Adesea asemenea itemi sunt eliminați din test.

Itemii la care toată lumea răspunde corect sunt folosiți în unele teste ca itemi de debut, pentru familiarizarea subiectului cu sarcina, și nu se iau în considerare la calculul scorului testului.

Deși au varianță mică, itemii ușori, la care aproape toate persoanele răspund corect, sunt incluși în testele prin care se încearcă să se depisteze absența unor deprinderi sau a unor cunoștințe elementare. De exemplu, cînd este testată competența profesională după un curs de instruire, este de așteptat ca subiecții să răspundă corect la majoritatea itemilor testului de cunoștințe, deci itemii să apară ca fiind ușori.

Variabilitatea scorurilor la un test format din itemi necorelați liniar între ei este maximă atunci cînd toți itemii au indicele de dificultate egal cu 0.5.

Unii autori recomandă păstrarea în test a itemilor al căror indice de dificultate este cuprins între .20 și .80. În funcție de scopul urmărit la aplicarea unui test se vor utiliza itemi mai dificili sau mai ușori.

Validitatea itemilor

Prin validitatea unui item înțelegem calitatea acestuia de a măsura ceea ce-și propune să măsoare testul din care itemul face parte (creativitatea, responsabilitatea, empatia etc.). Această definiție este o simplă extensiune a definiției clasice a validității testului.

Pentru a evalua validitatea unui item se alege un criteriu extern și se calculează coeficientul de corelație dintre scorurile

acestuia și scorurile itemului. Vor fi eliminați din test toți itemii pentru care coeficientul de corelație astfel calculat este nesemnificativ statistic la un prag fixat (de obicei $p=0.05$). Adesea în construcția testelor se utilizează unul dintre următoarele criterii :

- un alt test, consacrat, care măsoară același lucru ca și testul pe care îl proiectăm;
- cotarea paralelă, de către mai mulți specialiști, a fiecărei persoane examinate;
- performanța profesională a subiecților (în cazul testelor de cunoș-ințe sau de aptitudini speciale).

Formulele statistice utilizate pentru calculul coeficientului de corelație sunt dependente de modul de răspuns la itemii respectivi (itemi dihotomici sau continui etc).

Validitatea itemilor relativă la un criteriu extern nu trebuie confundată cu validitatea testului relativă la constructul măsurat (APA Standards, 1985).

Capacitatea de discriminare a unui item

Aprecierea capacității de discriminare a unui item se bazează pe compararea răspunsurilor obținute la item fie cu performanțele realizate de aceiași subiecți la criteriul utilizat pentru validarea itemului, fie cu scorurile testului din care face parte itemul. Se fac următoarele raționamente:

- Dacă un item este valid relativ la un criteriu extern, adică măsoară același lucru ca și criteriul, atunci este de așteptat ca majoritatea persoanelor care obțin performanțe mari la criteriu să răspundă la item corect, iar majoritatea celor care realizează performanțe scăzute la criteriu, să răspundă eronat.

- Dacă itemul măsoară același lucru ca și ceilalți itemi ai testului, atunci persoanele care răspund corect la itemul studiat răspund corect la majoritatea itemilor, deci obțin scoruri mari la test, în timp ce persoanele care greșesc la acest item răspund eronat la mulți itemi și realizează un scor scăzut la test.

Pentru ca un item să fie eficient, nu este suficient ca el să măsoare ceea ce se dorește să măsoare testul din care face parte (adică itemul să fie valid relativ la un criteriu sau să măsoare același lucru ca și ceilalți itemi), ci trebuie să fie capabil să *discrimineze* între subiecții "superiori" și cei "inferiori" în privința caracteristicii măsurate (adică între subiecții care obțin la criteriu, respectiv la test, performanțe superioare și cei care realizează performanțe scăzute). Capacitatea (puterea) de discriminare a unui item este studiată folosind: indicele de discriminare al unui item, corelația item-test și corelațiile inter-itemi.

Indicele de discriminare al unui item

Murphy și Davidshofer (1991) propun o metodă de măsurare a puterii de discriminare a unui item care utilizează două grupe extreme, alcătuite pe baza rezultatelor obținute de subiecți la un criteriu extern (sau la întregul test din care face parte itemul). Tehnica de lucru este simplă. Subiecții examinați cu testul analizat, care obțin rezultate extreme la criteriu (sau la întregul test), sunt structurați în două loturi:

- un lot al celor care au performanțe superioare (grupa bună);
- un lot cu acei subiecți care au performanțe inferioare (grupa slabă).

Murphy și Davidshofer (1991), citîndu-l pe E.E. Cureton, recomandă ca fiecare grupă să conțină între 25% și 33% din totalul celor cărora li s-a administrat testul.

Pentru fiecare item se calculează indicele de discriminare D, cu ajutorul formulei:

$$D=C_B/N_B-C_S/N_S$$

unde:

N_B, N_S = numărul de persoane în grupa bună, respectiv în grupa slabă (dacă testul a fost administrat cu limită de timp, atunci N_B și N_S vor reprezenta numărul de persoane care au parcurs itemul, în fiecare din cele două grupe);

C_B, C_S = numărul persoanelor care au rezolvat itemul corect, în grupa bună, respectiv în grupa slabă.

Deci, indicele de discriminare *D* este diferența a două frecvențe relative, mai precis este diferența indicilor de dificultate ai itemului în două grupe contrastante. El variază între -1 (cînd toți cei din grupa slabă au rezolvat itemul corect și toți cei din grupa bună au greșit) și +1 (cînd itemul a fost rezolvat corect de toți cei din grupa bună și greșit de toți cei din grupa slabă). Cu cît indicele de discriminare este mai ridicat, cu atît întrebarea/itemul reușește mai bine să diferențieze între cei buni și cei slabi.

Valoarea indicelui de discriminare *D* este dependentă de aceea a indicelui de dificultate, *p*. Cînd toți subiecții rezolvă corect itemul (p=1) sau cînd nimeni nu răspunde corect (p=0), itemul nu diferențiază între cei buni și cei slabi. Dacă valoarea *p* este apropiată de 0 sau de 1 capacitatea de discriminare a întrebării este redusă. Cel mai mare potențial pentru a fi buni discriminatori îl au itemii al căror indice de dificultate este apropiat de 0.5.

Exemplu (Murphy & Davidshofer, 1991). Un test format din 40 de itemi a fost analizat de mai mulți experți, care au apreciat că toți itemii măsoară același lucru. Testul a fost administrat unui lot de 100 studenți. 27 de studenți cu scoruri mari la test au format grupa bună, iar 27 de studenți cu scoruri mici la test au format grupa slabă. S-a calculat frecvența relativă a celor care au răspuns corect la fiecare item al testului, în fiecare din cele două grupe. Tabelul 4-2 prezintă rezultatele obținute la patru itemi ai testului.

Tabelul 4-2
Frecvența relativă a subiecților care au răspuns corect la primii 4 itemi ai testului

Item	Grupa bună	Grupa slabă	D
1	0.71	0.42	0.29
2	0.60	0.24	0.36
3	0.47	0.42	0.05
4	0.38	0.61	-0.23

Examinînd coloana *D* a tabelului 4-2, observăm că itemul 3 are o putere de discriminare slabă: el este aproape la fel de dificil pentru ambele grupe, deci nu diferențiază între subiecții buni și cei slabi. În ceea ce privește itemul 4, rezultatul analizei este contrar așteptărilor: el are o putere discriminativă mare în valoare absolută, dar cu semn negativ (itemul este ușor pentru subiecții cu performanțe slabe la test și dificil pentru cei cu performanțe bune). Este posibil ca acest item să fie formulat greșit și din acest motiv persoanele din grupa bună să fie derutate la alegerea răspunsului corect.

Unul din obiectivele urmărite la construcția testelor este să se formuleze itemi care, la fel ca itemii 1 și 2 ai testului analizat, să discrimineze corect între două loturi contrastante.

Corelația item-test.
O procedură mai simplă de evaluare a puterii de discriminare a unui item constă în calcularea coeficientului de corelație liniară dintre scorurile obținute la item și scorul total la test. Obișnuit, itemii sunt cotați dihotomic (1 pentru răspuns corect și 0 pentru răspuns greșit sau omis); în acest caz este utilizat coeficientul de corelație biserial.

Coeficientul de corelație item-test se interpretează la fel ca și indicele de discriminare al itemului (*D*). Un coeficient de corelație pozitiv, semnificativ statistic la un prag fixat (de obicei, p=0.05), ne spune că itemul diferențiază între cei cu performanțe mari la test și cei cu performanțe scăzute. O corelație item-test apropiată de zero, nesemnificativă statistic, ne arată că itemul nu face o discriminare între cei cu scoruri mari la test și cei cu scoruri mici. Corelația negativă semnifică faptul că între test și item există o neconcordanță: cei care rezolvă corect itemul obțin scoruri mici la test, iar cei care îl rezolvă greșit, au scoruri mari.
Sunt eficienți doar itemii pentru care coeficientul de corelație item-test este semnificativ statistic, la pragul fixat, și pozitiv.

Corelațiile inter-itemi
Calculul coeficienților de corelație liniară dintre itemii testului, luați doi cîte doi, ajută la înțelegerea validității și a puterii de discriminare pentru fiecare item.

Dacă toți itemii testului corelează slab între ei, înseamnă că majoritatea lor măsoară altceva decît ceea ce trebuie să măsoare și, în consecință, trebuie reformulați.

Atunci cînd se evidențiază grupuri de itemi, cu proprietatea că itemii fiecărui grup corelează puternic între ei, dar nu corelează semnificativ statistic cu itemii celorlalte grupuri, înseamnă că

fiecare grup măsoară un alt atribut. În asemenea cazuri nu se va studia puterea de discriminare a itemilor prin compararea scorurilor itemilor cu scorurile testului.

Curba caracteristică a unui item
Pentru studiul testelor psihologice care măsoară atribute specifice ale persoanei, precum inteligența, creativitatea etc., a fost dezvoltată *teoria răspunsului la itemi* (TRI) (sau teoria trăsăturii latente - *latent trait theory*) (Lord, 1980; Rasch, 1966).

Aceasta analizează calitățile fiecărui item prin intermediul *curbei caracteristice a itemului* (CCI) (Lord, 1980; De Gruijter & Van der Kamp, 1984; Albu & Pitariu, 1993).

Curba caracteristică a unui item dihotomic este graficul funcției care ia ca valoare, pentru fiecare valoare posibilă a atributului măsurat de test, $y \in \mathbf{R}$, probabilitatea ca o persoană care are nivelul y al atributului respectiv, să obțină scorul 1 la item, $P(y)$. Această reprezentare realizează o vizualizare a comportării itemului ca instrument de măsură. Cu ajutorul ei pot fi apreciate calitățile itemului (validitate, dificultate și putere de discriminare).

Întrucît se presupune că atributele măsurate de teste iau valori pe întreaga axă reală, curba caracteristică a unui item este graficul unei funcții P, definită pe axa reală, cu valori în intervalul [0,1]:

$$P : \mathbf{R} \to [0,1]$$

În cazul itemilor valizi, această curbă este nedescrescătoare. Într-adevăr, este normal ca, pe măsură ce crește nivelul atributului măsurat de test, să crească și probabilitatea de a se obține la item scorul 1. De exemplu, pentru un item dintr-un test de atenție, este de așteptat ca probabilitatea de rezolvare corectă a itemului să fie mai mare la persoanele care au un nivel ridicat al atenției decît la cele care au un nivel scăzut.

Teoria răspunsului la itemi deplasează centrul de greutate de la scorul testului la modul de rezolvare a itemului. Este astfel aprofundat mecanismul de răspuns, strategia cognitivă de elaborare a scorului la test.

În Figura 4-1 este ilustrată curba caracteristică a unui item valid conținut într-un test de inteligență. Ea este graficul funcției care descrie probabilitatea obținerii răspunsului corect la item în funcție de nivelul de dezvoltare al inteligenței. Itemul constă dintr-o întrebare cu alegere multiplă, care propune un răspuns corect și trei răspunsuri greșite. Toate răspunsurile sunt bine formulate, astfel încît nici unul dintre ele nu este mai atractiv sau mai puțin atractiv unei persoane cu un nivel foarte slab al inteligenței. Răspunzînd la întîmplare, aceasta va avea deci 25% șanse să dea răspunsul corect.

Din acest motiv, valorile funcției *P* reprezentată grafic în figura 4-1 pornesc de la probabilitatea de 0.25 și nu de la 0. Cu cît crește nivelul de dezvoltare al inteligenței, cu atît crește și probabilitatea de a răspunde corect la item, la niveluri foarte ridicate ale inteligenței probabilitatea fiind apropiată de valoarea 1. Alura crescătoare a funcției *P* ne sugerează că itemul măsoară corect inteligența.

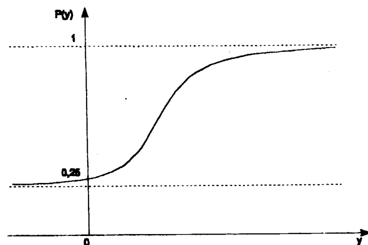
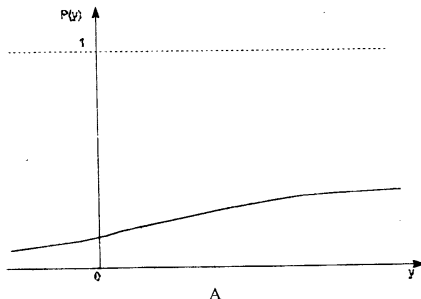
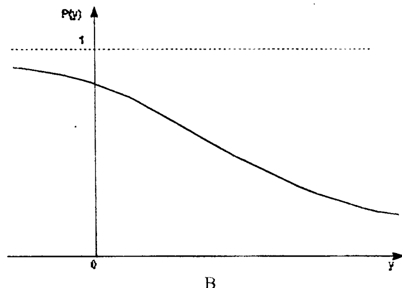


Figura 4-1 Curba caracteristică a unui item (CCI)

În practică putem întîlni diferite CCI, fiecare avînd o semnificație aparte. În Figura 4-2 sunt reprezentate CCI pentru doi itemi dintr-un test de inteligență.



A



B

Figura 4-2 A - CCI pentru un item valid, dificil și cu putere de discriminare mică
B - CCI pentru un item care nu este valid

Primul item este valid (funcția *P* este crescătoare), dar este foarte dificil (valorile funcției *P* sunt scăzute chiar și pentru niveluri înalte ale inteligenței) și are o putere de discriminare scăzută la toate nivelurile de inteligență (valorile funcției *P* variază foarte puțin, tangenta la CCI în orice punct este aproape paralelă cu axa Ox) (Figura 4-2 -A). Cel de-al doilea item nu este valid (funcția *P* este descrescătoare, deci probabilitatea de rezolvare corectă a itemului este mai mare la niveluri scăzute ale inteligenței decît la niveluri ridicate) (Figura 4-2 - B).

Cu ajutorul CCI se poate studia puterea de discriminare a itemului la fiecare nivel $y \in \mathbf{R}$ al atributului măsurat de item. Aceasta este indicată de panta tangentei la curba funcției *P*, în punctul *y*. Itemul discriminează puternic atunci cînd panta este mare și discriminează puțin cînd panta este apropiată de zero.

CCI poate fi utilizată și pentru a analiza gradul de dificultate al unui item. În Figura 4-3 sunt trasate curbele caracteristice pentru trei itemi despre care putem spune că au o putere de discriminare asemănătoare, la fiecare nivel $y \in \mathbf{R}$, dar a căror dificultate este diferită. Astfel, itemul *C* este cel mai dificil (valorile funcției *P* sunt cele mai scăzute), iar itemul *A* este cel mai ușor de rezolvat (funcția *P* corespunzătoare are valorile cele mai ridicate). Dificultatea este deci identificată după poziția ocupată de CCI pe verticală. Pentru itemii dificili, valorile funcției *P* pornesc din apropierea lui 0 și nu depășesc cu mult valoarea 0.5. Pentru itemii foarte ușori, funcția *P* ia valori mai mari decît 0.5, și se apropie de 1 atunci cînd y este foarte mare.

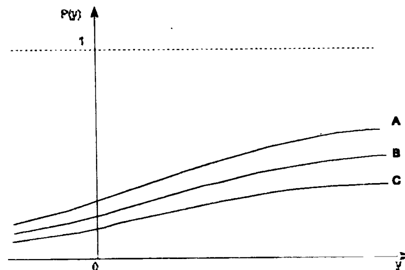


Figura. 4-3 Reprezentarea grafică a curbelor caracteristice pentru trei itemi de test cu dificultate diferită

Curba caracteristică a itemului și teoria răspunsului la itemi
Teoria răspunsului la itemi oferă o modalitate de analiză a relației dintre particularitățile individuale ale persoanei (aptitudini, cunoștințe etc.) și răspunsurile sale la itemii care intră în componența testelor prin care sunt măsurate aceste particularități.

Aplicarea teoriei răspunsului la itemi pentru un anumit test debutează cu formularea unui set de presupuneri referitoare la relația matematică dintre aptitudinile unei persoane și probabilitatea ca aceasta să răspundă într-un anumit fel (corect sau incorect, prin "da" sau "nu") la fiecare item al testului, la diverse niveluri ale atributului măsurat de test. Aceste presupuneri constituie o bază pentru alegerea formei analitice a funcțiilor P ale itemilor. De obicei se consideră că forma analitică a funcției P este aceeași pentru toți itemii unui test, ceea ce deosebește itemii unii de alții fiind doar parametrii care apar în expresia funcțiilor P și care se numesc parametrii itemului.

La construirea curbei caracteristice a unui item intervine un aspect deosebit de dificil: valorile atributului măsurat de item (aptitudini, cunoștințe sau priceperi) nu sunt și nici nu pot fi cunoscute, pentru nici o persoană. În teoria răspunsului la itemi sunt descriși, însă, unii algoritmi care permit ca, pe baza presupunerilor formulate în legătură cu itemii și a răspunsurilor date la test de către un grup de persoane să se estimeze atît parametrii itemilor testului, cît și valorile atributului măsurat de itemi, la toate persoanele examinate (Lord, 1980; De Gruijter & Van der Kamp, 1984).

Aplicații practice ale TRI
Aplicarea în practică a teoriei răspunsului la itemi este complexă, dar ea a fost mult ușurată în zilele noastre datorită utilizării calculatoarelor și a pachetelor de programe statistice. Se apreciază că, datorită informațiilor bogate pe care le dă despre calitățile

itemilor unui test, această teorie va deveni populară, în viitor ea deschizînd noi perspective analizei psihometrice a testelor.

TRI este utilă în cazul unor probleme care sunt dificil de rezolvat folosind mijloacele psihometrice tradiționale. Vom prezenta în continuare trei aplicații de acest fel ale TRI, și anume la:

- utilizarea răspunsurilor distorsionante pentru evaluarea aptitudinilor;
- utilizarea testelor adaptative (taylored testing);
- analiza și construirea testelor specializate.

Obținerea informațiilor prin analiza itemilor distorsionanți. Analiza de itemi tradițională privește răspunsul dat de un subiect la un item cu alegere multiplă ca fiind "corect" sau "greșit", fără a se interesa prea mult care din variantele greșite oferite ca răspunsuri alternative a fost aleasă și din ce motiv.

TRI pleacă de la ideea că orice răspuns dat de un subiect la un item al testului ne poate spune ceva despre cunoștințele sau aptitudinile persoanei respective. În acest context, se determină pentru fiecare variantă de răspuns propusă de item *curba caracteristică a răspunsului* (CCR). Asemănător curbei caracteristice a unui item, CCR indică, pentru fiecare nivel al atributului măsurat de item, $y \in \mathbf{R}$, probabilitatea ca o persoană avînd nivelul y să aleagă răspunsul respectiv. Dacă răspunsurile propuse de item sunt bine formulate, atunci CCR corespunzătoare răspunsului corect este crescătoare (din aceleași rațiuni ca și CCI), în timp ce curbele caracteristice ale răspunsurilor eronate sunt descrescătoare și apropiate între ele (la orice nivel al atributului măsurat, orice persoană care nu cunoaște răspunsul corect indică la întîmplare una dintre variantele propuse, avînd aceeași probabilitate să aleagă oricare dintre ele).

Figura 4-4 ilustrează graficele curbelor caracteristice ale răspunsului la un item cu alegere multiplă care propune un răspuns corect (*A*) și trei răspunsuri greșite (*B*, *C* și *D*). Despre variantele de răspuns *A* si *B* putem spune că ele sunt bine formulate: curbele caracteristice ale răspunsului sunt monotone (crescătoare pentru *A* și descrescătoare pentru *B*). Faptul că varianta de răspuns *C* are CCR crescătoare, apropiată de CCR pentru varianta *A* sugerează ipoteza că răspunsul *C* nu este bine formulat: el se aseamănă nepermis de mult cu răspunsul corect *A*. Varianta *D* este aleasă cu o probabilitate scăzută și se pare că nu are nici o legătură cu atributul măsurat de item (CCR este aproape paralelă cu axa Ox). În concluzie, se impune înlocuirea variantelor *C* și *D*.

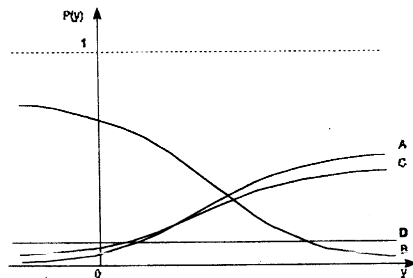


Figura 4-4 CCR pentru un item cu alegere multiplă, cu patru variante de răspuns

Teste adaptative

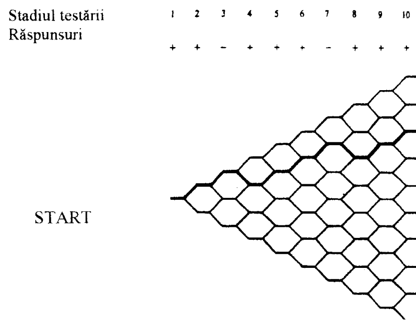
Un test de cunoștințe de fizică este compus dintr-un anumit număr de itemi. Unii dintre aceștia sunt mai grei, alții sunt de dificultate moderată iar alții sunt ușor de rezolvat. Persoanele cărora li se administrează testul în cauză diferă foarte mult din punctul de vedere al cunoștințelor de fizică pe care le posedă. Pentru cei mai bine pregătiți, itemii prea ușori, rezolvați cu ușurință de toți subiecții, sunt plictisitori și lungesc inutil timpul de administrare a testului, în timp ce pentru persoanele slab pregătite, itemii foarte dificili sunt obositori și, din acest motiv, prezența lor în test poate influența modul de rezolvare a celorlalți itemi. De la asemenea observații a apărut ideea creării de teste adaptative.

Testele adaptative permit identificarea nivelului de cunoștințe sau a nivelului aptitudinii măsurate pentru fiecare subiect, folosind un număr mic de itemi, care sunt selectați, dintr-o colecție existentă, chiar în timpul administrării testului, în funcție de răspunsurile date de persoana examinată. Deci, un test adaptativ diferă de la un subiect la altul.

Cînd se administrează un test adaptativ, fiecare item al testului este considerat un pas, la nivelul căruia se evaluează cunoștințele sau aptitudinile persoanei examinate. Pe baza rezultatului acestei evaluări se alege itemul care se administrează în continuare.

O sesiune de examinare cu un test adaptativ debutează de obicei cu un item de dificultate moderată. Dacă subiectul răspunde corect, se trece la un item mai dificil, în caz contrar se alege un item mai ușor. Procedura continuă în acest mod pînă cînd se poate formula o concluzie despre subiect.

Figura 4-5 ne prezintă schema constructivă a unui test adaptativ, a cărui colecție de itemi este alcătuită din 55 de itemi. Fiecare item al testului este reprezentat printr-un segment orizontal care constituie o latură a unui hexagon sau o latură comună la două hexagoane plasate unul deasupra celuilalt. Poziția pe verticală a segmentului este determinată de nivelul de dificultate al itemului. Cu cît itemul este mai greu, cu atît segmentul corespunzător este plasat mai sus.



Linia îngroșată ne indică traseul parcurs de examinat de-a lungul examinării (o liniuță- corespunde unui item de test).
Figura 4-5 Schema constructivă a unui test adaptativ.

Construirea testului adaptativ în timpul examinării unei persoane constă în alegerea a 10 itemi, deci a 10 segmente orizontale. Administrarea testului începe totdeauna cu itemul al cărui segment se află în partea stîngă a desenului, a cărui extremitate dreaptă coincide cu vîrful triunghiului format din hexagoane. Itemul administrat la fiecare pas corespunde unui segment care are extremitatea dreaptă comună cu vîrful unui hexagon. Dacă subiectul răspunde corect, atunci se va continua cu itemul corespunzător laturii orizontale superioare a acestui hexagon, în caz contrar se va alege itemul care corespunde laturii orizontale inferioare a hexagonului. În Figura 4-5 se pot identifica itemii care au fost administrați unei persoane. Aceștia corespund segmentelor orizontale ale liniei îngroșate. Poziția pe verticală a segmentului corespunzător ultimului item administrat vizualizează nivelul la care se află persoana examinată. În cazul ilustrat în Figura 4-5, subiectul se află deasupra nivelului mediu, dar nu atinge nivelul maxim posibil.

Construirea manuală a unui test adaptativ este foarte dificilă. Din acest motiv, teoria și practica testării adaptative s-au dezvoltat doar după ce a început să se extindă pe scară largă utilizarea calculatoarelor în testarea psihologică și în testarea cunoștințelor.

Examinarea cu ajutorul testelor adaptative face ca activitatea de testare să fie mult mai eficientă, economică și flexibilă (Landy, Shankster & Kohler, 1994; Schoonman, 1989).

Analiza de itemi pentru teste specializate

Testele de aptitudini speciale măsoră atribute sau constructe care stau la baza efectuării unor activități specifice. Vorbim astfel despre teste de aptitudini mecanice, teste de matematică sau pentru alte discipline etc. Construirea unor asemenea teste respectă, în general, următoarele trei condiții :

1. Variantele de răspunsuri incorecte trebuie să fie la fel de atractive pentru orice persoană lipsită total de atributul măsurat de test.
2. Indicele de dificultate trebuie să fie apropiat de valoarea 0.50.
3. Corelația item-test trebuie să fie pozitivă și semnificativă statistic la un prag fixat (de obicei $p=0.05$).

Unele tipuri de teste, cum sunt cele de screening sau testele în care scorul nu depinde de corectitudinea sau incorectitudinea răspunsului, constituie cazuri speciale. Construcția lor nu ia în considerare toate condițiile formulate mai sus.

Testele destinate screening-ului sunt acelea care se utilizează la decizii preliminare sau pentru a determina dacă o persoană deține sau nu un nivel minim de cunoștințe sau deprinderi care să-i permită să practice o anumită activitate. De pildă, pentru a urma școala de șoferi un candidat trebuie să posede un nivel minim de acuitate vizuală și capacitate de a percepe distanța (percepția adîncimii).

Desigur, prezența calităților amintite nu fac pe nimeni un conducător auto mai bun; absența lor este însă o contraindicație profesională cît se poate de clară. Testele de screening sunt utilizate și ca filtru în activitatea de selecție profesională, cînd numărul de candidați este în exces față de cel al locurilor disponibile. În acest caz se aplică un test de screening, prin care se face o triere grosieră a subiecților după un criteriu precis, fixat anterior împreună cu beneficiarul selecției (analiza CV-ului va permite respingerea celor care au o vîrstă peste baremul stabilit; bărbații pot fi respinși de la un post de secretară; pot fi respinși cei cu valori anormale ale tensiunii arteriale etc.).

Unele teste nu cer răspunsuri corecte sau greșite ci doar acordul sau dezacordul cu o afirmație sau alta (criterion-keyed test). Scorul la un astfel de test este interpretat pe baza comparării sale cu scorurile obținute de un grup de referință. De exemplu, *Inventarul Psihologic California* (CPI) cere subiectului să precizeze, la fiecare din cei 462 de itemi ai săi, dacă este de acord sau nu cu o anumită afirmație. Profilul psihologic al unei persoane, obținut pe baza scorurilor scalelor din CPI, se poate compara cu profilul mediu al unui grup profesional, școlar etc.

La ambele tipuri de probe pe care le-am amintit, pentru construirea și analizarea testelor, se pot aplica procedurile specifice TRI.

Să presupunem că dorim să selecționăm o secretară pentru o companie. Un item ideal proiectat în scopul efectuării unui prim filtru al candidaților va avea o curbă caracteristică similară cu aceea din Figura 4-6. Curba are o creștere bruscă la nivelul pragului de admisibilitate fixat de companie.

Subiecții care au un nivel al atributului măsurat de item mai scăzut decît acest prag au șanse foarte reduse să rezolve itemul, în timp ce candidații cu un nivel mai ridicat au șanse mari. Un test scurt, construit după principiile expuse, va fi eficient pentru un astfel de filtru/screening preliminar, prin care se elimină din start o parte din candidații la post.

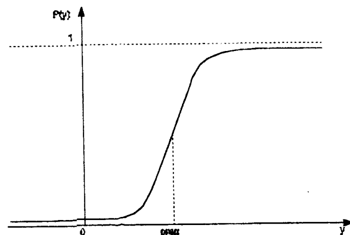


Figura 4-6. Reprezentarea grafică a unui item ideal utilizat într-o acțiune de screening.

În ceea ce privește testele la care nu există răspunsuri bune sau greșite, procedura de analiză de itemi va urma o cale aparte. Se vor colecta răspunsurile la itemi pentru cîteva criterii diferite (de exemplu, pentru mai multe grupe profesionale) și se va determina curba caracteristică a fiecărui item, pentru fiecare criteriu ales.

Figura 4-7 ilustrează curbele caracteristice obținute la un item al CPI (267. *Prefer să povestesc decît să ascult.*) de către trei grupe profesionale: profesori, ofițeri și ingineri. Acest item intră în componența scalei Do (Dominanță), fiind cotat cu un punct răspunsul *Adevărat* și cu zero puncte răspunsul Fals.

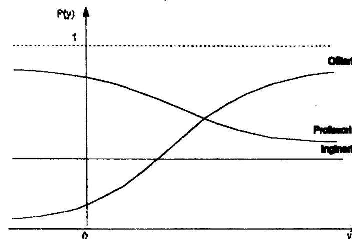


Figura 4-7. CCI pentru itemul 267 din CPI, la trei categorii profesionale (profesori, ofițeri și ingineri) .

Din Figura 4-7 se observă că itemul este valid pentru scala Do doar în cazul grupului de ofițeri. Răspunsurile date de ingineri nu au nici o legătură cu nivelul de dominanță al acestora. În schimb, pentru profesori itemul nu este potrivit scalei Do: răspunsul Adevărat este mai frecvent la profesorii care au un nivel scăzut de dominanță decît la cei cu un nivel înalt, deci primii vor obține un punct în plus la scala Do.

Putem formula cîteva concluzii cu privire la abordarea modernă a analizei de itemi, prin tehnicile exemplificate în acest capitol. Astfel, este un lucru cît se poate de limpede că analiza de itemi modernă oferă un cadru experimental și informațional superior.

Măsurările bazate pe TRI sunt printre cele mai complete privind detectarea categoriilor de distorsiuni care pot caracteriza itemii din care este construit un test.

De exemplu, în Figura 4-8 sunt ilustrate curbele caracteristice ale unui item dintr-un test de atenție, pentru bărbați și pentru femei. În ceea ce privește capacitatea itemului de a discrimina între persoanele cu niveluri înalte și cele cu niveluri scăzute ale atenției, se poate vedea că itemul este un bun discriminator pentru femei și unul slab pentru bărbați.

Itemul apare mai dificil pentru bărbati decît pentru femei (CCI corespunzătoare bărbaților se află sub CCI corespunzătoare femeilor).

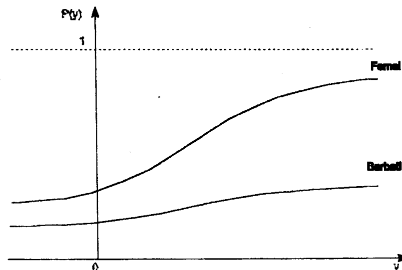


Figura 4-8. Un item a cărui curbă caracteristică este dependentă de sex.

TRI permite nu numai să se identifice distorsiunile provocate de itemi, ci și să se prevadă consecințele acestora în diverse contexte de aplicare a testului. Astfel, de exemplu, utilizarea unui test alcătuit din itemi asemănători celui reprezentat în Figura 4-8 poate avea următoarele consecințe:

- bărbații vor obține la test scoruri mai mici decît femeile;
- scorurile testului vor fi aceleași sau foarte apropiate ca valoare pentru persoane de ambele sexe care au un nivel scăzut al atenției, chiar dacă ele diferă mult unele de altele în privința acestui nivel;
- toate femeile cu un nivel ridicat al atenției, chiar dacă acest nivel diferă, vor obține scoruri identice sau apropiate ca valoare;
- pentru femei se vor obține cu frecvențe mari doar două categorii de scoruri: mici (pentru niveluri ale atenției

mai mici decît un prag corespunzător punctului în care panta tangentei la CCI are valoarea maximă) și mari (pentru niveluri ale atenției mai mari decît acest prag);

- pentru bărbați scorurile testului nu vor fi polarizate la cele două extremități (mici și mari), așa cum se întîmplă la femei.

BIBLIOGRAFIE

Albu, M., Pitariu, H. (1993). Proiectarea testelor de cunoștințe și examenul asistat de calculator. Editura Casa Cărții de Știință, Cluj-Napoca.

Albu, M. (1995). Construirea și utilizarea testelor psihologice. GAR 4936.

Allport, G.W. (1970). Structure et dévelopment de la personalité. Delacaux et Niestlé, Neuchâtel.

Anastasi, A. (1954). Psychological Testing, MacMillan Publishing Co., Inc., New York.

Anastasi, A. (1976), Fields of Applied Psychology, McGraw-Hill, Inc., N.Y.

Anastasi, A. (1982). Psychological Testing. McMillan Publishing Co., Inc., N.Y.

Arthur, A.Z. (1969). Diagnostic testing and the new alternatives. Psychological Bulletin, 72, 3, 183-192.

Bacher, F. (1981). Les enquétes en psychologie, Université de Lille III.

Blum, M. L., Naylor, J. C. (1968). Industrial Psychology. Its Theoretical and Social Foundations, Harper&Row, Publishers.

Bordens, K.S., Abbott, B.B. (1991). Research Design and Methods. A Process Approach. Mayfield Publishing Company: London.

Brogden, H. E., Taylor, E. K. (1949). The Dollar Criterion - Applying the Cost Accouting Concept to Criterion Construction, Personnel Psychology.

Brownstein, S.C., Weiner, M. (1955). How to Prepare for College Entrance Examinations. Barron's Educational Series, Inc., NY.

Carroll, J.B., Maxwell, S.E. (1979). Individual differences in cognitive abilities. Annual Review of Psychology, 30, 603-640.

Cascio, W.F. (1991). Applied Psychology in Personnel Management. (4th ed.). Prentice-Hall, Inc.,: New Jersey.

Corsini, R. J. (Ed.) (1994). Encyclopedia of Psychology, John Wiley & Sons.

Cronbach, L. J. (1943). On Estimates of Test Reliability, The Journal of Educational Psychology, 34, 485-494.

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16, 297-334.

Cronbach, L. J. (1966). Essentials of Psychological Testing, A Harper International Edition.

Cronbach, L.J., Gleser, G.C. (1965). Psychological tests and personnel decisions. (2rd ed.). Urbana, Ill, University of Illinois Press.

Cureton, E. (1955). Validity. In : Lindquist, E. F. (ed.), Educational Measurement, American Council on Education, Washington, D.C.

Cronbach, L.R. (1966). Essentials of Psychological Testing. Harper & Row, N.Y.

De Gruijter, D. N. M., Van der Kamp, L. J. Th.(1984). Statistical Models in Psychological and Educational Testing, Swets & Zeitlinger, Lisse.

Dunnette, M. (1966). Personnel Selection and Placement. Wadsworth Publishing, Co., Inc. Belmont, California.

Endler, N. S., Magnusson, D. (1976). Interactional Psychology and Personality, Hemisphere Publishing Corporation, Washington, D.C.

Eysenck, H.J. (1963). The measurement of motivation. Scientific American, 208 (5), 130-140.

Fisseni, H. J. (1990). Lehrbuch der psychologischen Diagnostik, Verlag für Psychologie. Dr. C. J. Hogrefe, Gottingen.

French, W.L. (1990). Human Resources Management. Second Edition. Houghton Mifflin Company, Boston.

Gilmer, H.B von (1971). Industrial and Organizational Psychology, McGraw-Hill Book Company.

Graziano, A. M., Raulin, M. L.(1993). Research Methods. A Process of Inquiry, Harper Collins College Publishers.

Grondlund, N. E. (1967). Measurement and Evaluation in Teaching, The MacMillan Company, New York.

Guilford, J. P.(1954). Psychometric Methods. McGraw-Hill Book Company,Inc., New York and London.

Guilford, J. P. (1965). Fundamental Statistics in Psychology and Education, McGraw-Hill Book Company.

Guion, R.M. (1965). Personnel Testing. McGraw-Hill Book Company, N.Y.

Guion, R.M. (1974). Open a new window: Validities and values in psychological measurement. American Psychologist, 29, 287-296.

Guion, R. M. (1986). Changing Views for Personnel Selection Research, Lucrare prezentată la al 21-lea Congres Internațional de Psihologie, Ierusalim, 15 iulie 1986.

Gulliksen, H. (1950). Theory of Mental Tests. N.Y.: Wiley.

Hammond, S. (1995). Using Psychometric Tests, În: Breakwell, G. M., Hammond, S., Fife-Schaw, C. (eds.), Research Methods in Psychology, Sage Publications.

Kendall, K. G. (1948). Rank Correlation Methods, Griffin, London.

Kerlinger, F.N. (1986). Foundations of Behavioral Research. (3rd edition). Arcourt Brace Jovanovic College Publisers: New York.

Kuder, G.F., Richardson, M.W. (1937). The theory of the estimation of test reliability. Psychometrika, 2, 151-160.

Kulcsar, T. (1980). Lecții practice de psihodiagnostic. Univ. Babeș-Bolyai, Cluj-Napoca.

Landy, F. L.(1985). Psychology of Work Behavior, The Dorsey Press Chicago, Illinois.

Landy, F. L. (1986). Stamp Collecting Versus Science. Validation as Hypothesis Testing, American Psychologist, Vol. 41, 11, 1183-1192.

Landy, F.J.., Shankster, L.J., Kohler, S.S. (1994). Personnel selection and placement. Annual Review of Psychology, 45, 261-296.

Lefter, V., Manolescu, A. (1995). Managementul resurselor umane. EDP, R.A., București.

Lord, F. M. (1980). Applications of Item Response Theory to Practical Testing Problems, Lawrence Erlbaum Associates, Publishers, Hillsdale, New Jersey.

Lord, F.M., Novick, M.R. (1968). Statistical Theories of Mental Test Scores, Addison-Wesley Publishing Company.

Murphy, K. R. (1987). Principles of Psychological Testing, Prentice Hall, Enlewood Cliffs, N. J.

Murphy, K.R., Davidshofer, Ch. O.(1987; 1991). Psychological Testing: Principles and Applications. Prentice Hall, Englewood Cliffs, N.J.

Nevo, B. (1993). Face Validity Revised 1., În: Nevo, B., Jäger, R. S.(ed.). Educational and Psychological Testing: The Test Taker's Outlook, Hogrefe&Huber Publishers.

Novick, M.R., Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. Psychometrika, 32, 1-13.

Pavelcu, V. (1974). Drama psihologiei. EDP, București.

Păun, G. (1987). Paradoxurile clasamentelor, Editura Științifică și Enciclopedică, București.

Pitariu, H. (1975). Testul Domino 48. Testul Domino 70. în: îndrumător psihodiagnostic. Vol II, Universitatea "Babeș-Bolyai"

Pitariu, H. (1978). Psihologie industrială - Lecții. Universitatea Babeș-Bolyai, Cluj-Napoca.

Pitariu, H., O aplicație a teoriei deciziei la examinările psihologice, În: Eficiența examenului psihologic în selecția personalului de la siguranța circulației rutiere, Prima consfătuire a psihologilor din transporturi, Suceava, 19-21 noiembrie 1976.

Pitariu, H. (1994). Managementul resurselor umane. Măsurarea performanțelor profesionale, Editura ALL, București.

Rasch, G. (1966). An item analysis which takes individual differences into account. British Journal of Mathematical ans Statistical Psychology, 19, 49-57.

Roe, R. (1995). Developments in Eastern Europe and Work and Organizational Psychology. în: C.L. Cooper, I.T. Robertson (ed). International Review of Industrial and Organizational Psychology. Vol 10, N.Y.: John Wiley & Sons, 275-349.

Roșca, M. (1972). Metode de psihodiagnostic, Editura Didactică și Pedagogică, București.

Rulon, P.J. (1939). A simplified procedure for determining the reliability of a test by split-halves. Harward Educational Rev., 9, 99-103.

Schoonman, W. (1989). An Applied Study on Computerized Adaptive Testing. Swets&Zeitlinger B.V., Amsterdam/Lisse.

Silva, F. (1993). Psychometric Foundations and Behavioral Assessment. Sage Publications, Inc.

Stevens, S.S. (1951). Mathematics, measurement, and psychophysies. în S.S. Stevens (ed). Handbook of Experimental Psychology. New York: Wiley, 1-49.

The US Department of Labor (1988). Occupational Outlook Handbook, VGM Career Horizons, Lincolnwood, Illinois.

Thorndike, R.L. (1949). Personnel Selection. John Wiley & Sons, Inc.

Thorndike, R.L., Haagen, E. (1961). Measurement and Evaluation in Psychology and Education. John Wiley & Sons, Inc.

Traub, R. E. (1994). Reliability for the Social Sciences. Theory and Applications, Vol. 3, Sage Publications, 1994.

Wesman, A.G. (1952). Reliability and confidence. (Test Service Bulletin No.56). N.Y.: The Psychological Corporation.

*** Grand dictionnaire de la psychologie (1994). Larousse, Paris.

*** Standards for Educational & Psychological Tests (1974; 1985). Published by the American Psychological Association, Inc.

*** Dictionary of Education (1959). New York, Toronto, London.

STATI-GODEA
MARIANA

Tipărit la *Editura* Mesagerul
B-dul Eroilor nr. 2
3400 CLUJ-NAPOCA
ROMÂNIA

ERATĂ

Pagina 13:

Paragraful al doilea se continuă cu rândurile 16 şi 17 (formula Cns)
Ultimul paragraf se şterge, fiind o repetiţie

Pagina 41:

Ultimul rând, formula corectă este:

$m_{nou}(a)=p+qm(a)$, **pentru orice** $a \in A$

Pagina 130:

Rândul 5 de jos, în loc de (x + 5)2=(y + 1)3, se va citi: $(x + 5)^2=(y + 1)^3$
Rândul 1 de jos, în loc de III. X=(y + 1)3/2 - 5, se va citi: III. $x = (y + 1)^{3/2}-5$

COLECȚIA UNIVERSITARIA NR. 2

În Colecția Universitaria au apărut:

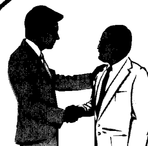
- Vasile Surd: "Geografia dezvoltării și a decalajelor economice contemporane"

În Colecția Universitaria vor apărea:

- Ioan Aurel Pop: *Istoria Transilvaniei medievale: de la etnogeneza românilor până la Mihai Viteazul*
- Militon Frențiu și col.: *Elaborarea algoritmilor*
- Militon Frențiu și col.: *Programarea Pascal*
- Rodica Frățilă: *Studiul mărfurilor alimentare*
- Rodica Frățilă și Marinela Ghereș: *Calitatea mărfurilor alimentare*
- Vasile Stanciu: *Muzica bisericească corală*
- Ecaterina Popa: *Elemente de sintaxă engleză*
- George Postelnicu: *Macroeconomie*

I.S.B.N. 973-9261-19-1

UNIVERSITARIA



HORIA PITARIU
MONICA ALBU

Psihologia personalului I. Măsurarea și interpretarea diferențelor individuale

EDITURA
PRESA UNIVERSITARĂ CLUJEANĂ

UNIVERSITARIA